THIRD ANNUAL

Ecological Quality Assurance

Workshop

PROCEEDINGS

Canada Centre for Inland Waters

Burlington, Ontario

APRIL 1990

91-03071/

PROCEEDINGS OF THE

THIRD ANNUAL ECOLOGICAL QUALITY ASSURANCE WORKSHOP

Meeting Coordinated by:
John Lawrence, Ph.D., and
Robert Bisson, Ph.D.
National Water Research Institute
Burlington, Ontario

Proceedings Coordinated by:
Donald R. Hart, Ph.D.
Beak Consultants Limited
Brampton, Ontario

Sponsored by:

Environment Canada, Conservation and Protection
National Water Research Institute
Canada Centre for Inland Waters
Burlington, Ontario, and

U.S. Environmental Protection Agency
Environmental Monitoring Systems Laboratory
Las Vegas, Nevada

Held on April 24-26, 1990 at:

Canada Centre for Inland Waters
Burlington, Ontario

EXECUTIVE SUMMARY

It is becoming increasingly important in the environmental sciences to demonstrate the quality of any data produced. Data from contaminated site investigations must often withstand close scrutiny in court. Moreover, decisions on abatement and remedial actions often hinge on the quality of available data.

The question, "How good are the data?" is often asked with some implicit data use in mind. It is becoming increasingly clear that a more pertinent questions is, "Are the data good enough to meeting my needs?" An answer to this question requires explicit definition of intended data uses and data quality needs. If these parameters are well defined prior to data collection it is usually possible to design the data collection effort to ensure that the data are good enough, and to confirm this by monitoring data quality throughout the study.

The principles of quality assurance (QA) and the concept of integrated quality management, have been applied for some time in the manufacturing sector. They have recently been recognized in the environmental sciences, particularly with regard to chemical measurement systems, and are beginning to be adapted to biological measurement systems.

Over 80 people from Canada, the United States and Europe, representing government, university and private industry sectors, participated in this Third Annual Ecological Quality Assurance Workshop, which was held at the Canada Centre for Inland Waters in Burlington, Ontario, April 24-26, 1990. The purpose of the workshop was to provide a forum for interdisciplinary exchange of ideas and resolution of issues associated with QA in ecological studies, and thereby to promote the extension of QA principles into the planning and implementation of such studies.

Following on recommendations from the Second Annual Quality Assurance Workshop in 1989 in Las Vegas, Nevada, this third workshop focussed on QA for biological measurement systems, and for trace organic chemical monitoring. The structure included a plenary session (day 1) in which invited papers were presented on various aspects of ecological QA experience, followed by 6 concurrent workgroup sessions (day 2) dealing with different ecological subject areas. In a final plenary session (day 3) workgroup leaders or rapporteurs presented a summary of group activities, conclusions and recommendations.

Invited Papers

In total, 11 invited papers were presented. Dean Neptune (EPA, Quality Assurance Management Staff) began with an overview of the data quality objective (DQO) process, emphasizing the iterative nature of the process, and the need for 'up-front' and ongoing communication between the data suppliers and users. Don Hart (Beak Consultants Limited) presented an overview of biological QA guidelines for aquatic environmental protection, currently under development by Environment Canada, emphasizing general principles and checklists for program evaluation. Jim Pollard (Lockheed Engineering and Sciences Co.) addressed the use of the chemistry template in design of biological QA for ecology, emphasizing identification of control checkpoints in the data generating process, and corresponding performance or data quality measures. Bob Breckenridge (Idaho National Engineering Laboratory) discussed general QA issues associated with terrestrial ecosystem monitoring, emphasizing the dynamic nature of ecosystems, and consideration of natural variability in project design, site selection and sampling.

A number of authors related their experience with QA in particular ecological monitoring programs. Merete Johannessen (Norwegian Institute for Water Research) described an international European acid precipitation monitoring program, and intercalibration of the different analytical and site scoring methods used. Bill Burkman

(NSI Technology Services Corp.) described the joint U.S. - Canada sugar maple decline project, and illustrated the positive impact of annual crew training and evaluation sessions on precision of biological measurement. John Innes (U.K. Forestry Commission) described an international European forest response study, some of the major method differences between countries, and the improved reproducibility associated with crew training and evaluation. Dan McKenzie (EPA Environmental Research Laboratory (ERL)) described the new Environmental Monitoring and Assessment Program (EMAP) Inland Aquatic Program and the QA issues now being addressed at the design stage. John Paul (EPA - ERL) described the EMAP Near Coastal Program, its data quality objectives and the design alternatives currently being considered to meet these objectives.

Several authors described QA aspects of environmental contaminant monitoring programs. Richard Turle (Canadian Wildlife Service) described the use of biological reference materials (RMs) for organochlorines in estimating analytical precision and accuracy of waterfowl residue measurements. Bill Strachan (National Water Research Institute) described the mass balance modelling framework of a Canadian precipitation monitoring program, and the development of data quality objectives for this research project.

Workgroup Sessions

Workshop participants were divided into one of 6 workgroups, each group addressing a different ecological subject area, and associated QA issues and approaches.

Workgroup 1 discussed QA issues related to laboratory assessment of aquatic ecosystems. The group spent a good deal of time discussing lab certification. This issue appears to be of considerable importance to the environmental compliance programs in both Canada and the United States. The group concluded that, if good performance standards or criteria are developed for each method and results of performance were

included with data reports, a self-governing system would work. Other issues that were of major concern to the group were staff training, and reference toxicants or other toxicity test performance criteria. There was a realization that present reference toxicant testing is insufficient because animal performance is not segregated from analyst performance. Training is imperative and must be emphasized more in QA plans. The group identified a need to establish a standard protocol for methods development, which would include development of performance criteria. A consensus set of method performance criteria could reduce the length of time spent by government entities in defining methods developed for use in the regulatory community. Hypothesis setting and defining objectives for a study were also important considerations. It was noted that QA in the absence of a hypothesis or objective is non-functional.

Workgroup 2 discussed QA issues related to field assessment of aquatic ecosystems. The group began with a general discussion of issues to be addressed in QA planning, then focussed on a hypothetical biological monitoring project near an industrial discharge, to illustrate how these issues could be addressed. Participants spent some time working through the DQO process for this project. They noted the iterative nature of the process as project objectives and procedures were alternately refined and eventually matched to achieve an adequate study design, based on a logic statement for each decision-relevant variable. Major potential sources of error were identified along with QC elements to measure and control these components of total uncertainty within acceptable limits.

Workgroup 3 discussed QA issues associated with terrestrial ecosystem monitoring. The group defined a logical progression of issues to be considered in QA planning, including problem definition, variable prioritization, method selection, method documentation, crew training, data evaluation and validation, and use of data quality information. There was general agreement that data quality should be planned for and organized during the development of a program. It was noted that QA personnel need to be involved with overall project personnel, not as a separate group. This finding also

applied to statistical and data management personnel. A number of unnresolved issues were identified, including:

- how to assess accuracy of field measurements.
- how to deal correctly with questionable data,
- how to assess comparability of data between years,
- how to keep field crew personnel involved with and enthusiastic about the project, and
- how to remove the separation between project management and QA (i.e. What is the best process to combine these aspects of a study).

Workgroup 4 discussed QA issues related to controlled field and laboratory experiments in terrestrial ecosystems. The group started with a general discussion of issues to be addressed in QA planning. Participants identified several critical issues including:

- how to ensure that QA is integrated into the planning and development of scientific research, and
- how to append QA/QC data to the study results in a tractable manner.

The group then proceeded to consider two hypothetical case studies - a field study to develop fertilization criteria for a controlled experiment, and a controlled experiment to evaluate fertilization effects on growth and development under acidified conditions. Major potential sources of error were identified, along with appropriate QA activities to measure and/or control these components of uncertainty.

Workgroup 5 discussed quality management for monitoring trace toxic organics in air and precipitation. The group concluded that data quality management for atmospheric sampling networks should include detailed documentation of procedures and all observations, including those from the field, during sample transport and in the laboratory. Training of field personnel was also considered highly important. Auditing of conformation to established procedures and timely inspection of data were considered essential to provide feedback to the operation of the network, including all aspects of sample collection, analysis and data storage. A variety of measures to track data quality were presented.

Workgroup 6 discussed quality management for monitoring trace toxic organics in an aquatic ecosystem. The group considered a hypothetical crisis situation as a rationale for further monitoring (dead fish found near a dredge spoil disposal operation, some with elevated tissue residues). They briefly addressed the environmental questions and related quality management issues that would be expected to arise, and then developed approaches to these issues as part of a monitoring QA plan. Sampling design considered compositing of specimens, to obtain sufficient sample volume, and a sufficient number of composite samples to meet uncertainty objectives in comparison of impacted and reference sites, with attention to possible fish size and species effects. Bounds of analytical performance were defined to keep total uncertainty within acceptable limits, and reporting requirements were developed to include appropriate quality documentation.

Conclusions and Recommendations

Based on the invited papers and workgroup discussions, it was generally agreed that QA in environmental chemistry, which is reasonably well understood, provides a useful template for QA in environmental biology and ecology, although some adjustments in approach are required for ecological implementation of the common QA principles. A number of common themes emerged:

- the concept of total quality management, with integration of QA functions into every stage of project planning, implementation and reporting,

- the notion of quality assurance as sound scientific method to ensure that useful results are obtained,

- an emphasis on "up-front" planning with preliminary studies as required to permit development of an adequate project plan,

- an emphasis on statistical study design tailored to achievement of specific project and data quality objectives,

- the need for identification of quality check points in the data generation process, and development of related performance measures,

- the need for staff training and continual performance assessment of individuals, methods and results,

- the need for consistent QA terminology.

Suggestions for future directions of the workshop focussed on these themes, with continuing emphasis on QA approaches for biological measurement systems. Specific suggested topics included taxonomic QA, statistical design, biological QC samples for performance assessment, specimen banking, specific QA for trend evaluation, and method comparability needs in air monitoring.

NOTICE

Although this workshop was jointly sponsored by Environment Canada and the U.S. Environmental Protection Agency, the proceedings have not been subjected to formal review by either agency, and the opinions of individual authors do not necessarily reflect agency views. No official endorsement should be inferred.

ABSTRACT

This document presents manuscripts of invited papers, and reports of workgroup discussions, conclusions and recommendations arising from the Third Annual Ecological Quality Assurance Workshop, which was held at the Canada Centre for Inland Waters in Burlington, Ontario, April 24-26, 1990.

The purpose of this workshop was to provide a forum for interdisciplinary exchange of ideas and resolution of issues associated with quality assurance (QA) in ecological studies, and thereby to promote the extension of QA principals into the planning and implementation of ecological studies.

Over 80 people from Canada, the United States, and Europe attended the workshop representing government, university and private industry sectors, and many different scientific disciplines. This diverse mix of participants permitted sharing of experience and ideas on QA approaches across disciplines. There were 11 invited papers presented, and 6 concurrent workgroup sessions dealing with different ecological subject areas. Workgroup topics included:

Aquatic Ecosystems

- QA Issues Related to Laboratory Assessment
- QA Issues Related to Field Assessment

Terrestrial Ecosystems

- QA Issues Associated with Monitoring
- QA Issues Related to Controlled Experiments

Trace Toxic Organic Contaminants

- Quality Management for Monitoring in Air and Precipitation
- Quality Management for Monitoring in an Aquatic Ecosystem

Based on the invited papers and workgroup discussions, it was generally agreed that QA in environmental chemistry, which is reasonably well understood, provides a useful template for QA in environmental biology and ecology, although some adjustments in approach are required for ecological implementation of the common QA principles. A number of common themes emerged:

- the concept of total quality management with integration of QA functions into every stage of project planning, implementation, and reporting,
- the notion of quality assurance as sound scientific method to ensure that useful results are obtained,
- an emphasis on "up-front" planning, with preliminary studies as required to permit development of an adequate project plan,
- an emphasis on statistical study design, tailored to achievement of specific project and data quality objectives,
- the need for identification of quality checkpoints in the data generation process, and development of related performance measures,
- the need for staff training and continual performance assessment of individuals, methods and results,
- the need for consistent QA terminology.

Suggestions for future directions of the workshop focussed on these themes, with continuing emphasis on QA approaches for biological measurement systems. Specific

suggested topics included taxonomic QA, statistical design, biological QC samples for performance assessment, specimen banking, specific QA for trend evaluation, and method comparability needs in air monitoring.

ACKNOWLEDGEMENTS

The success of this workshop was due to the diligent efforts and contributions of many individuals, including the Workshop Organizing Committee, invited speakers, workgroup leaders and rapporteurs. The organizing committee consisted of Jim Lazorchak (EMSL, Cincinnati), Dennis McMullen (Technology Applications Inc.), Craig Palmer (NSI Technology Services Corp.), Keijo Aspila, Robert Bisson and John Lawrence (NWRI, Burlington), and Don Hart (Beak Consultants Limited). Mr. Dan Michaels (Research Triangle Institute) provided assistance to several workgroups in derivation of data quality objectives. Ms. Nancy Wentworth (EPA, Washington) kindly permitted use and reproduction of a draft QA glossary.

CONTENTS

Executive Summary i
Notice ix
Abstract xi
Acknowledgements xv
Introduction 1

PLENARY SESSION

"Applying the Data Quality Objective (DQO) Process to Research Operations" 5
Dean Neptune, Quality Assurance Management Staff,
U.S. Environmental Protection Agency, Washington, D.C.

"QA Guidelines for Biology in Aquatic Environmental Protection" 23
Don Hart, Beak Consultants Limited, Brampton, ON

"Biological Quality Assurance for Ecological Monitoring: Can We Use Chemistry as a Template?" 43
Jim Pollard, Lockheed Engineering and Sciences Co.,
Las Vegas, NV

"Intercalibration in the Framework of an International Monitoring Program" 55
Merete Johannessen, Norwegian Institute for Water Research,
Oslo, Norway.

"The Role of Quality Assurance in Conducting Terrestrial Ecosystem Monitoring" 73
Robert Breckenridge, Center for Environmental Monitoring and Assessment, Idaho National Engineering Laboratory, Idaho Falls, ID

"Quality Assurance Aspects of the Joint USA - Canada Sugar Maple Decline Project" 83
Bill Burkman, NSI Technology Services Corporation, Radnor, PA

"The Role of Quality Assurance in the Assessment of Forest Condition in Europe" 99
John Innes, Forestry Commission, Surrey, England

"QA Aspects of the EMAP Inland Aquatic Program" 113
Dan McKenzie, Environmental Research Laboratory, U.S. Environmental Protection Agency, Corvallis, OR

"Quality Assurance Aspects of EPA's Near Coastal Ecological Status and Trends Program" 121
John Paul, Environmental Research Laboratory, U.S. Environmental Protection Agency, Narragansett, RI

"In-House Reference Materials as a Means to QA: The CWS Experience" 131
Richard Turle, Canadian Wildlife Service, Environment Canada, Hull, PQ

"Trace Analysis of Atmospheric Precipitation Samples" 143
Bill Strachan, National Water Research Institute, Environment Canada, Burlington, ON

WORKGROUP SESSIONS

"Quality Assurance Issues Related to Laboratory Assessment of Aquatic Ecosystems" Workgroup 1 151

"Quality Assurance Issues Related to Field Assessment of Aquatic Ecosystems" Workgroup 2 165

"Quality Assurance Issues Associated with Terrestrial Ecosystem Monitoring" Workgroup 3 185

"Quality Assurance Related to Controlled Field and Laboratory Experiments in Terrestrial Ecosystems" Workgroup 4 193

"Quality Management for Monitoring Trace Toxic Organic Contaminants in Air and Precipitation" Workgroup 5 203

"Quality Management for Monitoring Trace Toxic Organic Contaminants in an Aquatic Ecosystem" Workgroup 6 209

Appendix A: Questionnaire and Summary

Appendix B: Quality Assurance Glossary

Appendix C: List of Participants

INTRODUCTION

Over 80 people from Canada, the United States, and Europe, representing government, university and private industry sectors, participated in the Third Annual Ecological Quality Assurance Workshop, which was held at the Canada Centre for Inland Waters in Burlington, Ontario, April 24-26, 1990. Hosted by Environment Canada's National Water Research Institute (NWRI), the workshop was co-sponsored by NWRI and the U.S. Environmental Protection Agency's Environmental Monitoring Systems Laboratory.

The purpose of this workshop was to provide a forum for interdisciplinary exchange of ideas and resolution of issues associated with quality assurance (QA) in ecological studies, and thereby to promote the extension of QA principles into the planning and implementation of ecological studies.

Following on recommendations from the Second Annual Quality Assurance Workshop in 1989 in Las Vegas, Nevada, this third workshop focussed on QA for biological measurement systems, and for trace organic chemical monitoring. The structure included a plenary session (day 1) in which 11 invited papers were presented on various aspects of ecological QA experience, followed by 6 concurrent workgroup sessions (day 2) dealing with different ecological subject areas:

Aquatic Ecosystems

- QA Issues Related to Laboratory Assessment
- QA Issues Related to Field Assessment

Terrestrial Ecosystems

- QA Issues Associated with Monitoring
- QA Issues Related to Controlled Experiments

Trace Toxic Organic Contaminaints

- o Quality Management for Monitoring in Air and Precipitation
- o Quality Management for Monitoring in an Aquatic Ecosystem

The format of each workgroup session was flexible, with some initial planning by the group leaders and rapporteurs, and with considerable input from participants on arrival as to QA/QC issues considered. Some groups adopted a generic approach to the issues, while others considered a hypothetical case study and developed the framework of an appropriate QA plan.

Workgroup leaders or rapporteurs presented a summary of group activities, conclusions and recommendations in a final plenary session (day 3), with brief open discussion following each of the six presentations.

This proceedings document is organized in two sections. Section 1 includes the text of invited papers, while Section 2 includes the workgroup reports. A summary of questionnaires completed by participants evaluating the meeting is presented in Appendix A, a proposed quality assurance glossary is presented in Appendix B, and a list of participants is presented in Appendix C.

PLENARY SESSION

Invited Papers

APPLYING THE DATA QUALITY OBJECTIVE (DQO) PROCESS TO RESEARCH OPERATIONS

Dean Neptune and Stan Blacker
Quality Assurance Management Staff
U.S. Environmental Protection Agency
Washington, DC 20460

Abstract

The Data Quality Objective process is an efficient way of getting the right people together to PLAN how to collect, in a cost-effective way, the right data needed to answer the key research question. These three core components -- right people, right data, and key question -- are essential. When planners of data operations are sensitive to each core component, and make an effort to address each properly, extraordinary savings in time, resources, and effort can result.

Unfortunately, the usual approach to gathering data has not always addressed these core components. In the absence of a more structured process, it is too tempting to start collecting data, without sufficient planning. Periodically one would pause from collecting data to review "progress" -- as measured by the quantity of data accumulating. And, one would assess whether the data actually answered the question, only after the money has been spent.

In a research arena, planning is all the more difficult, since one does not know what data are needed, or what measurements are feasible, until a significant commitment has been made to exploratory data gathering. We are sensitive to this issue, and have built into the DQO process a role for pilot studies, which themselves are capable of being optimized according to a DQO logic. These pilots stand in a phased relationship with the planning for the main data gathering effort -- the output from a pilot study serves as input for designing the next phase of the survey.

One aspect of research planning that makes a research DQO easier to prepare is the fact that researchers already carryout key elements of the DQO process. Their strong tradition of employing statistical experimental design aids in the development of a DQO.

This paper describes the basics of applying the DQO process to research planning and explains how this sort of attention during the planning of the design confers major benefits in time and cost-savings.

Introduction

The Data Quality Objective process is a structured approach to research planning. The process assures that key questions in planning research are asked of the right people, at the right time. When the research is limited in scope and participation, the DQO process reduces to the researcher addressing the following material by him/herself:

Defining the question or problem to be resolved,

Defining all the important variables that could affect the resolution of the problem,

Selecting those variables that need to be measured,

Establishing the manner in which these selected variables are to be assembled to answer the question (the logic statement), and

Determining the extent of uncertainty that the researcher is willing to accept in the answer and still conclude that an informative resolution of the question has been attained.

When the research is complicated, the researcher cannot carry out these steps alone. The research TEAM becomes involved. Each member of the team has an important question or problem that he/she wants to address. Sometimes, a researcher focuses only on his/her individual interest without understanding the broader question that needs resolution. Communicating the broader question to the team can be difficult. The team, needs to function as a team and appreciate that the project has broader questions needing attention.

The Data Quality Objective process reduces the individual researcher's desire to look at only his/her piece of the whole, and engages the research team in the necessary team dialogue to produce the research plan that addresses the broader question.

How Does the Data Quality Objective Process Work?

Defining the Problem

The first step in the DQO process is to define the problem. This is easy to say, but is generally one of the hardest steps to carry out. The difficulty arises because the research team usually has only a general, verbal idea of the problem they want to address. The verbal idea does not necessarily translate directly into questions that can be UNAMBIGUOUSLY resolved from a feasible data collection. Because the problems may be first conceived in fuzzy terms, researchers may be motivated to go out into the field and begin collecting that data that they routinely are comfortable collecting. The

general and unfocused data that are collected then might become the basis for narrowing and focusing the problem to be investigated. A second, or third, etc. data collection program ensues, with the researcher and team iterating between establishing a focused problem to be answered and collecting environmental data to address the problem. The "I'll know it when I see it" perspective results. This approach works perfectly well, but it is expensive and time consuming. Often one does not have the luxury or the time and resources for this sort of leisurely iteration of surveys.

With a more structured, focused approach, much, if not all, of the initial exploratory data collection effort could be replaced with careful "upfront planning". Enough knowledge and data are almost always available to establish and focus the problem, provided the right people are brought together to think about the problem in the right way.

In our approach to research planning, verbal statements of general problem are narrowed into written, specific, operational questions that can be answered UNAMBIGUOUSLY from defined sorts of data, BEFORE sending the team into the field. In this fashion, by the time researchers are sent into the field, we are sure that the data which they collect will constitute an answer.

Defining the Important Variables

Once the "focused" tentative question or problem needing resolution is defined, in writing, the researcher and team NEXT needs to define the important variables that affect the question or problem. There might be a large number of variables that could be relevant. Here, reliance on the skill of experienced and thoughtful researchers is required. They need to identify the smallest set of variables that is sufficient to answer the written question. In selecting this subset of variables, it is preferable to choose that subset of variables that are more readily measured.

Assembling the Key Variables into the Logic Statement

The next key step in the DQO process is to assemble these selected variables into a precise statement of the problem. This statement shows how the variables will be measured and combined to answer the question.

With such a statement, the researcher and team can revisit the initial problem and determine whether the research project defined by this logic statement really does address the problem to be pursued. If it doesn't, this shows that something is wrong either with the basic perception of the problem, or with the means chosen to pursue it. Or, the researchers may discover that their question was a good one, but that there is no practical way to answer it with the required resolution, within the constraint of the available resources. These eventualities are well worth discovering before an expensive program is underway in the field.

Acceptable Level of Uncertainty

The FINAL step in the DQO process is to ask the researcher and team to define the level of uncertainty that they are willing to accept in the answer and still conclude that they have an acceptable answer to their problem. This step establishes the design parameters for collecting and analyzing the data. These design parameters allow the technical staff to determine where to sample, how many samples to collect, how to analyze the samples, how good the analyses need to be, and how to assemble the data and present the desired result. It is this step that provides the information needed to perform the technical design for the field and laboratory efforts within the established uncertainty limits.

Upon the completion of this quantitative uncertainty statement, the researcher and team can determine whether they can achieve the desired data and quality. Two potential situations might arise. First, sufficient resources may not be available to collect the required data. Or, the underlying methodology may not be sufficiently advanced to achieve these data quality requirements.

The researcher and team have to choose between adding more resources or easing the data quality requirements or taking the time to develop new methodology. Having a precise and quantitative statement, empowers the researcher and team to know that they may have choices to make.

An additional thought is worth pointing out at this step in the DQO process. It is common to put a data gathering program in motion before the statistical rules for manipulating the expected data to get a "result" have been defined. Without knowing how to manipulate the expected data, the researcher and team cannot establish the likelihood that a proposed design will or will not prove adequate to the task. Our goal in use of the DQO process for research planning is to avoid surprises. By defining, in advance, the exact methodology that will be employed to turn data into a result, it becomes a routine matter to assess whether a proposed design will provide the results needed to answer the research question.

Summary

The Data Quality Objective process is a straight forward technique for researchers to use to answer three simple questions:

What is the precise research question to be investigated?

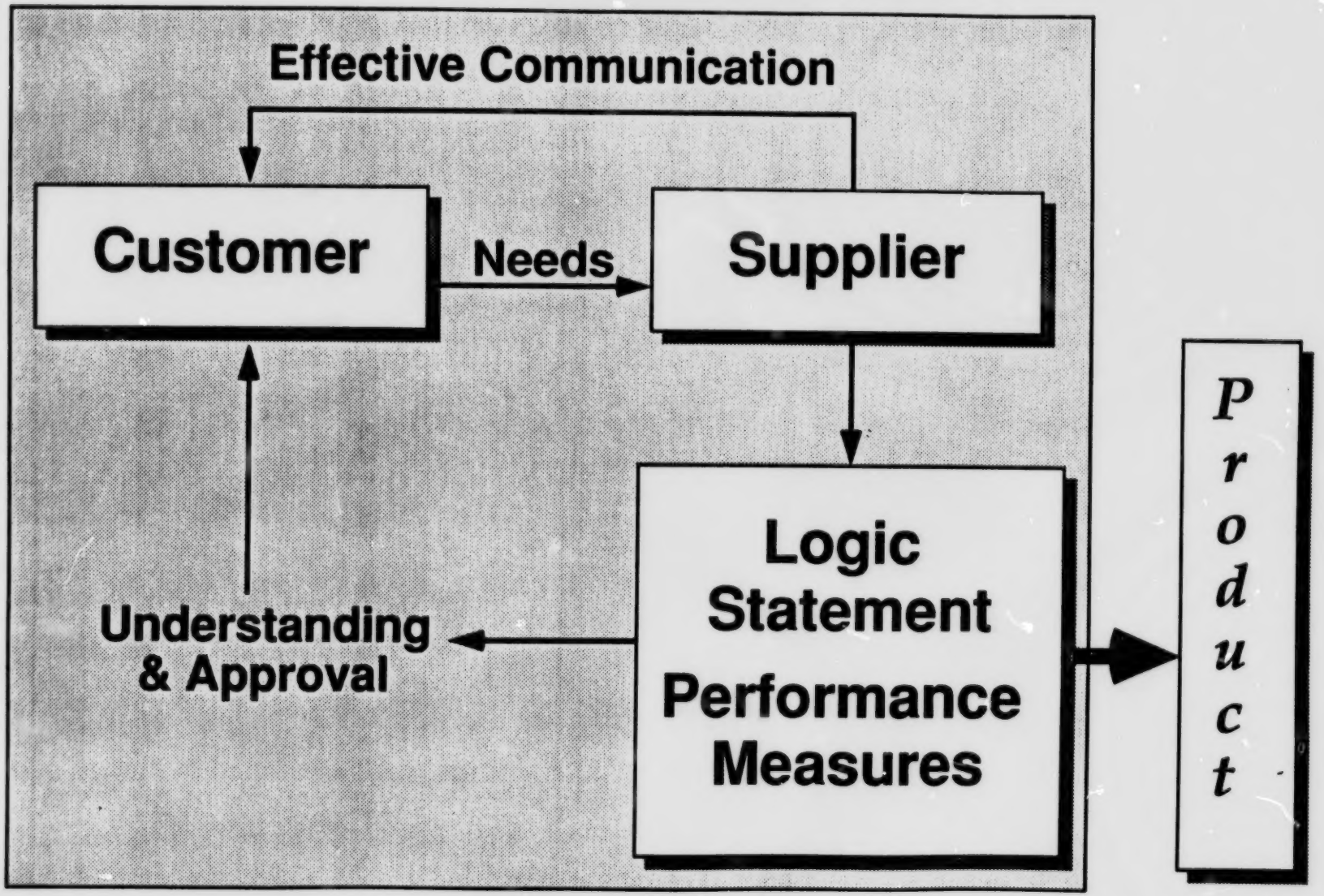
What are the important variables and associated acceptable level of uncertainty for which data need to be collected in order to answer the question posed?

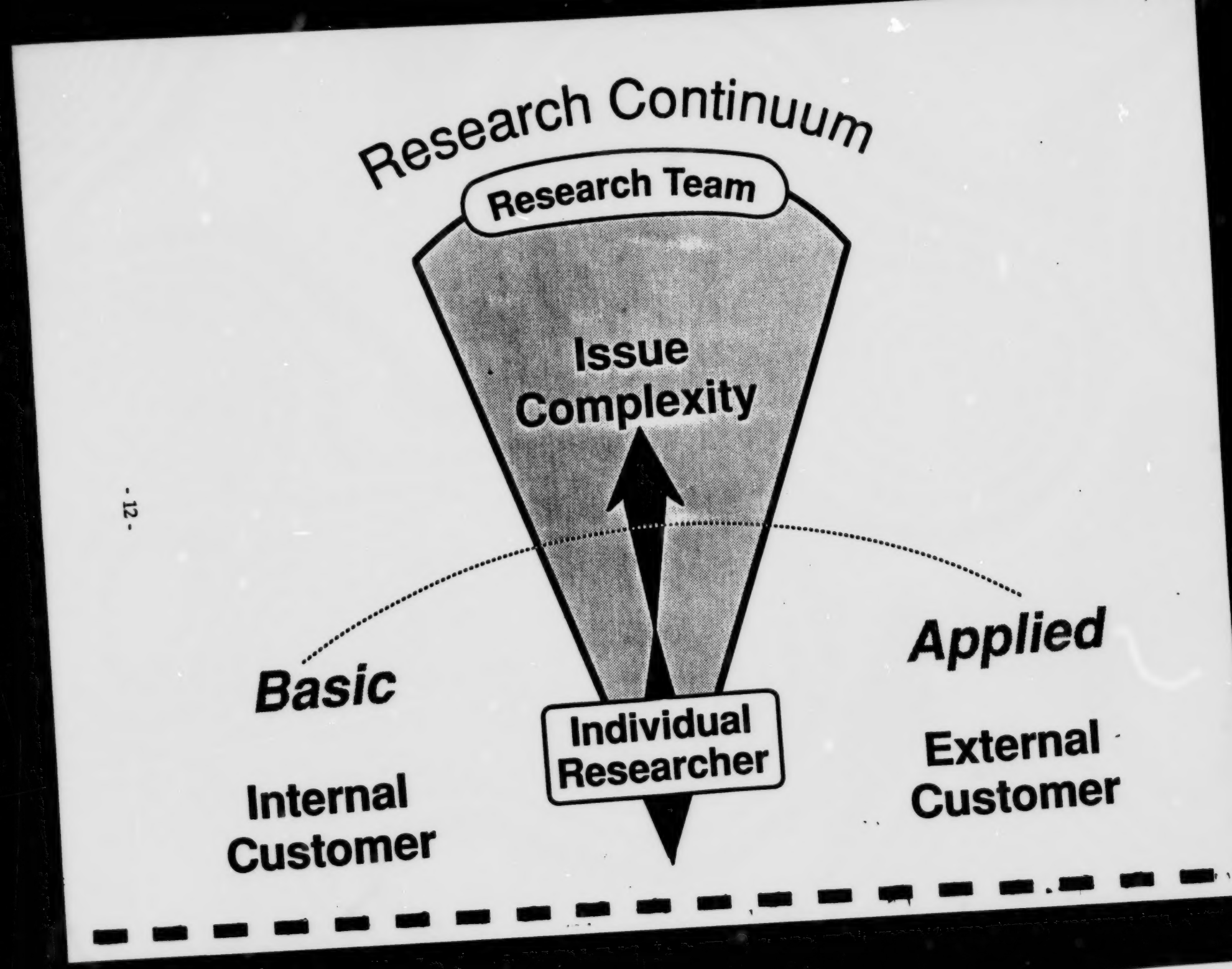
Can the needed data be collected with the available methodology and within the allocated budget?

It is often not easy to assemble the information in the correct form to answer these questions. The DQO process facilitates assembly and assessment of this information by providing a context in which the key players are brought together to consider the essential issues in a guided, step by step fashion. This facilitation occurs because the RIGHT people are being asked the RIGHT questions in order to pose the RIGHT statement of the problem to be addressed. All these RIGHTs add up to cost-effective experimental design. There is nothing unique about the DQO process. It is merely a carefully organized approach to carrying out thoughtful experimental design.

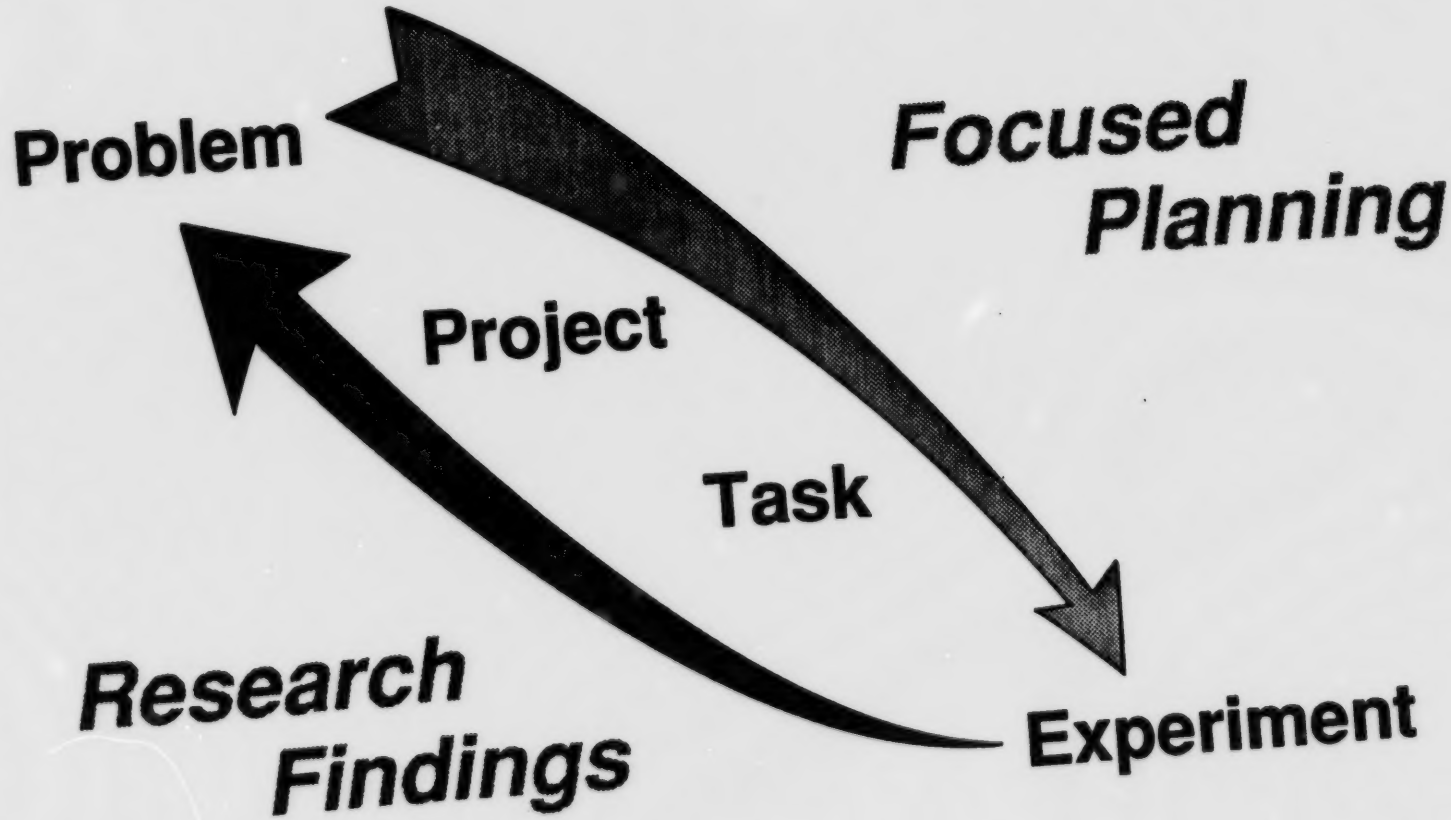
To successfully carry out the DQO process, the researcher and team have to be willing to invest in careful "upfront planning". There is no substitute. Although this planning does take time and resources, the leverage achieved by knowing what can be achieved before the data collection phase begins pays big dividends. Any researcher who is short on planning and fast on data collection, risks the embarassment of running out of resources before the question has been answered. The DQO process substantially reduces this risk.

Communication



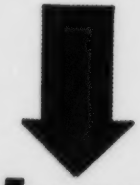

Research Continuum
Research Team
Issue Complexity
Basic
Applied
Individual Researcher
Internal Customer
External Customer

Research Hierarchy



Research Products

Basic

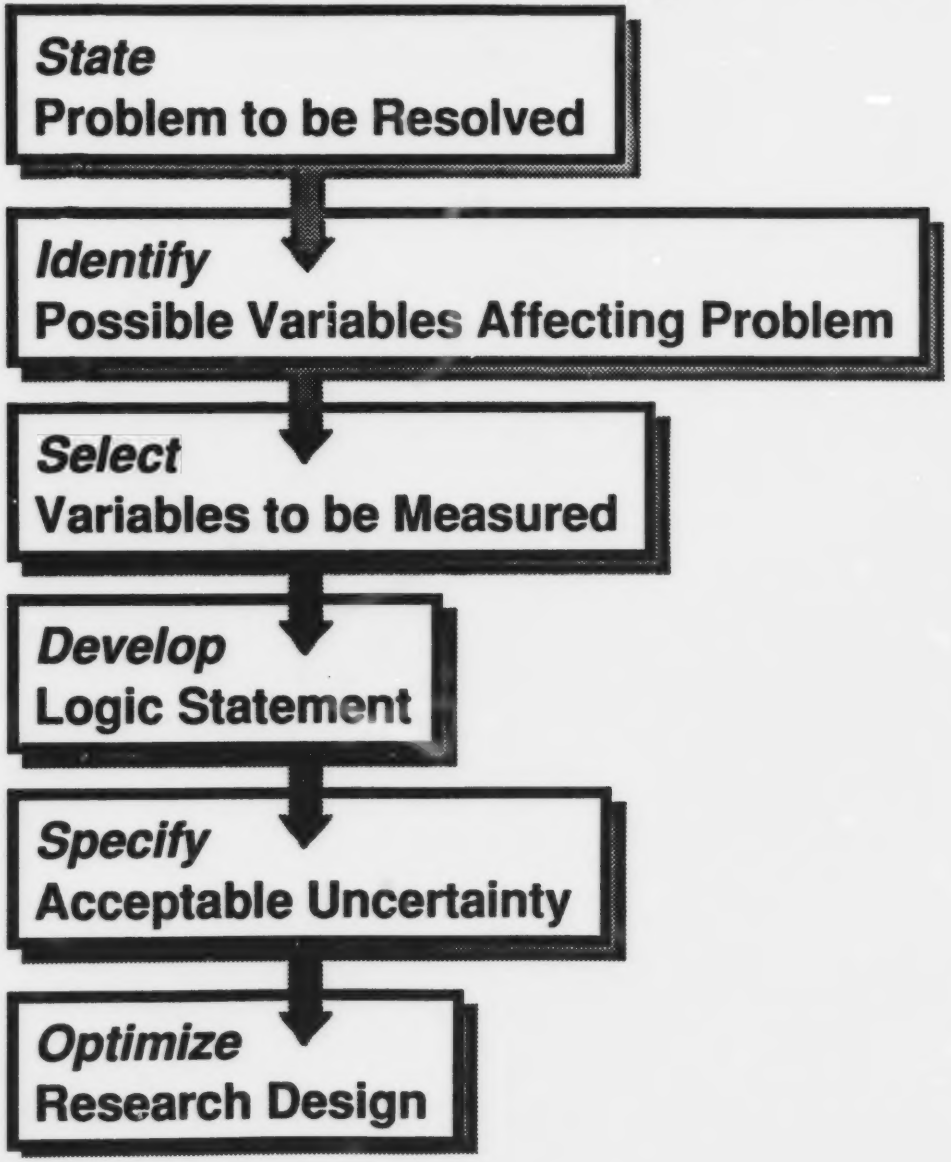


Test Hypothesis
to
further science

Applied

Answer Questions
to
determine appropriate actions

DQO Research Logic Flow



State the Problem to be Resolved

- What is known about the problem?
- How reliable is historical information?
- What remains to be found out?

Narrowly focused statement of problem

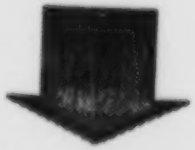
Identify Possible Variables Affecting Problem

- ID ***all*** important variables that might affect the question or problem
  - which variables are easier & more difficult to measure or control?

List all variables that are relevant & information that helps focus the list

Select Variables to be Measured

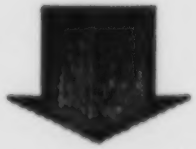
- Of these important variables, which are critical, useful or just nice to know?
- Is there more than one set of measurements that could be adequate to address the problem?



Carefully focused list of critical measurements

Develop Logic Statement

- How will variables be combined to provide a result that answers the question?
- Can you state in "if-then" terms how data will be used?
- Have you confirmed that all variables were identified & critical?



Concise statement or algorithm that unambiguously shows how data will be used to address the question

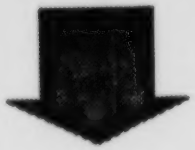
Specify Acceptable Uncertainty

- What are qualitative & quantitative consequences associated with incorrect answers?
- Are false (+) or false (-) errors of greater concern?
- What specific situation, if present, would you most want to control?

Qualitative statements that are translated into quantitative statements of the level of <u>uncertainty</u> the customer will accept in answers to the question

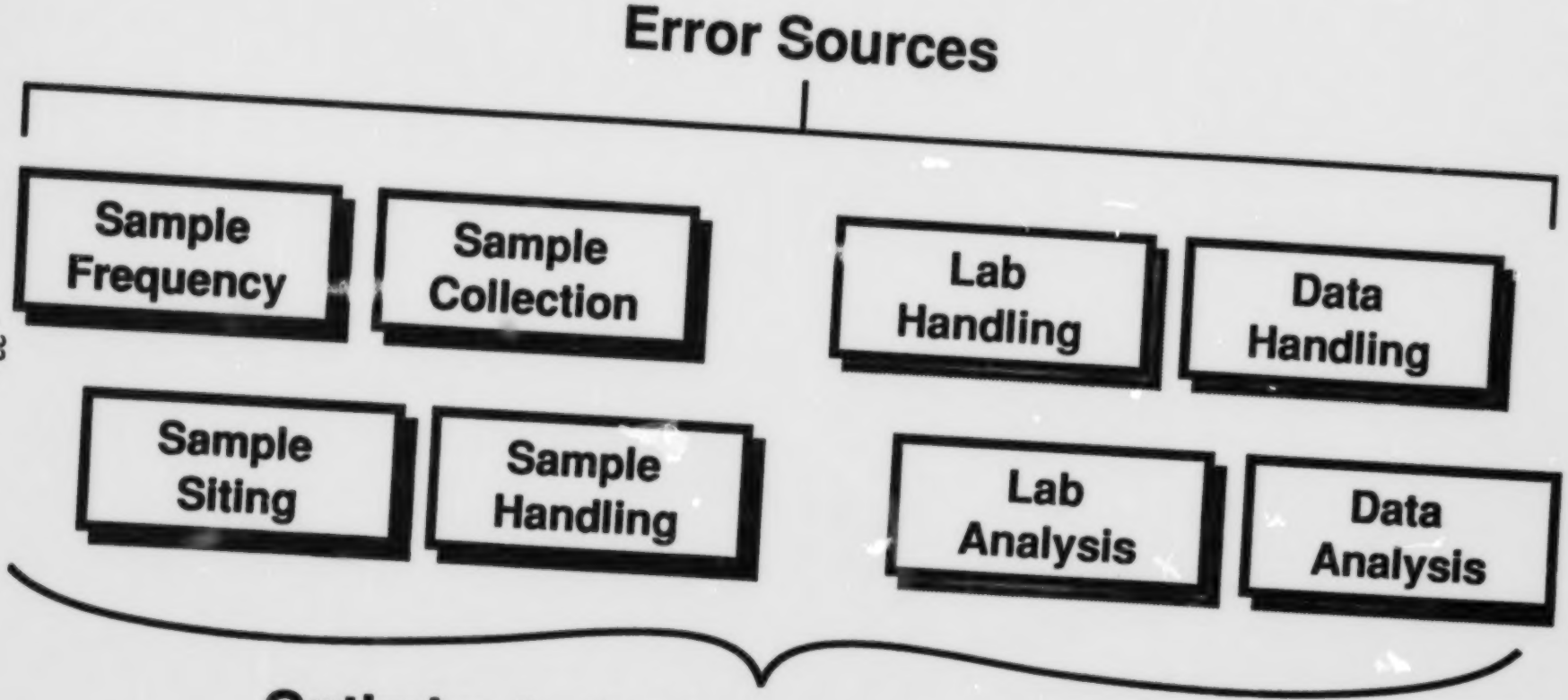
Optimize Research Design

- Can historical data be used?
- What design tools would lower variance in results for lower cost?
- How many samples needed?
- How will sample locations be determined?
- What sampling & analytical method will provide data of the desired quality?



The lowest cost design that will achieve desired limits on uncertainty

Design Optimization



Optimize individual errors to achieve total desired error at lowest cost

QA GUIDELINES FOR BIOLOGY IN AQUATIC ENVIRONMENTAL PROTECTION

Donald R. Hart
Beak Consultants Limited
14 Abacus Road
Brampton, Ontario, L6T 5B7

Abstract

Environment Canada is attempting to develop general guidelines on the design and implementation of quality assurance programs for biological measurement systems. The guideline document is intended to assist both government and private biological laboratories in establishing their own QA/QC programs, and to promote some degree of consistency in approach. It is recognized that biological data of known quality, assessed in terms of common data quality measures, will be increasingly needed with expanding use of biological testing in Canadian environmental protection.

The document describes the basic elements of a QA/QC program, based on principles of good laboratory practice, and then goes on to describe application of these principles in specific biological subject areas. Chapters are included on aquatic toxicity testing, microbiological assessment, aquatic community assessment and paleolimnology. Examples are given of data quality measures and data acceptance criteria, however, it is emphasized that data quality objectives should be tailored to intended data uses.

A QA checklist for toxicity laboratories is provided to encourage attention to specific aspects of laboratory operations, such as animal care and acclimation, temperature/light/ventilation control, water supply, food supply, dose selection, dose verification, experimental controls, test replication, LC50 calculation, reference toxicant testing, and toxicity control charting. A need for initiation of interlaboratory studies and follow-up investigations is identified.

The QA emphasis in aquatic community studies is on ensuring and quantifying the consistency and accuracy of sample collection and processing. Method inconsistencies associated with taxonomy, sampling seasons, sampling devices, sieve sizes, preservation and sorting techniques, often limit the suitability of data for temporal or spatial comparision. Sample replication is recommended to permit estimation of total precision, while split sample and reference collection exchanges may be used to quantify interlaboratory comparability. A check list is provided.

Introduction

Recent Canadian federal and provincial legislative initiatives provide the policy framework for expanding use of biological testing and assessment in environmental protection (e.g., Canadian Environmental Protection Act, Ontario Municipal Industrial Strategy for Abatement). Environment Canada has recognized the need for an increased level of effort in biological QA/QC if biological data are to be useful in a regulatory context. In 1988, it organized the Alliston Workshop (Day et al., 1988) to address issues related to use of biological methods in aquatic environmental protection. Participants concluded that general QA/QC guidelines for aquatic biological testing and assessment were needed.

In 1989, Beak Consultants Limited was commissioned to develop such a document, to reflect a consensus of the Canadian regulatory community on approaches to aquatic biological QA/QC in the field and the laboratory. The document is intended to provide guidance to both government and private institutions in the design and implementation of appropriate QA/QC programs. We are now at a third draft stage, and approaching a consensus. This paper will present some of the ideas that have emerged.

The guidance document emphasizes the principles of quality assurance, and focuses on particular problems that are encountered in application of these principles to biological measurement systems. The fundamental principles are the same, in biology, chemistry or even manufacturing. However, some of the potential sources of error are different. We attempt to identify these, and to suggest appropriate quality assessment and control techniques.

We have not written a methods manual, although the need for greater standardization in biological methodology is recognized (Day et al., 1988), along with the need for better documentation of methods and test conditions. Separate initiatives are underway to develop and publish Environment Canada protocols for specific toxicity tests, and these include QA/QC requirements. However, we are further from consensus as to detailed methodology for biological assessment in the field.

The scope of the guidance document now includes aquatic toxicity testing, microbiological assessment, aquatic community assessment and paleolimnology. Today, I intend to address aquatic toxicity and field assessment in general, but would like to begin at the beginning with a brief discussion of basic principles.

Elements of a QA/QC Program

The basic elements of a good QA/QC program can be packaged for presentation in many different ways. I use the following framework:

- o responsibilities and authorities - for institution and project;

- o sampling procedures - including sampling design;
- o analytical procedures - including biological analysis;
- o QA/QC procedures - packaged alone, or with test methods; and
- o documentation procedures - for storage, retrieval, revision and reporting of written methods, test results and supporting data.

Integration of project planning, sampling, analytical, QA/QC and documentation activities is the hallmark of any good measurement system. Principles of Good Laboratory Practice (GLP) cover the same territory, and are equally applicable in laboratory and field (e.g., OECD, 1981; ASTM, 1988).

A recent OECD (1989) draft council decision, which Canada has signed, calls for member countries to establish national GLP compliance monitoring procedures, designate authorities to monitor compliance, and require test facilities to issue a declaration of compliance with each study report. The scope of the program (i.e., chemical, biological) is open to definition.

Responsibilities and authorities (R&A's) define the organizational structure of the measurement system and associated QA/QC activities. A clearly defined and generally understood organizational structure with individual assignment of QA/QC functions, reporting channels and authority for specific corrective actions when established performance limits are not satisfied, is the backbone of the QA/QC program. This structure should be defined for each institution; and for each specific project. Projects may involve several institutions/laboratories whose roles and interactions must be defined and acknowledged. Key elements of R&A documentation include:

- o a Quality Management Plan - QA functions, reporting channels;
- o a Quality Assurance Policy - financial commitment to quality;
- o Quality Assurance Plans - for institution and project; and
- o a data certification system - with signatures and dates.

The QA policy statement is an essential inclusion in the Quality Management Plan. It should be signed by a Director with budgetary authority, in order to ensure allocation of resources for quality assurance, and to provide assurance of support to all staff in placing "quality first". It is difficult to overemphasize the importance of this short statement; it is the driving force behind the QA program.

Quality Management Plans may be developed for single institutions or laboratories (e.g., MOE, 1985) or for interagency programs (e.g., LRTAP, 1986; IJC, 1988). They are particularly critical in an interagency context since interagency structure tends otherwise to be loosely defined or non-existent. Similarly, Quality Assurance Plans may be developed for an institution or a project. Preparation guidelines have been provided by the EPA Quality Management Staff (QAMS, 1980).

Table 1 lists a few key elements of sampling, analytical and documentation procedures that require attention in development of a QA program, and documentation in Quality Assurance Plans and/or Standard Operating Procedures. The SOP document usually contains more technical detail on specific methods, and can be referenced in any Quality Assurance Plan.

I do not intend to elaborate here on each element, except to emphasize the importance of sampling procedures in general, and sampling design in particular, since these are often overlooked. Sampling error is usually a large component, and often the dominant component, of uncertainty in environmental measurements. With some preliminary sampling effort to estimate these components of uncertainty, as well as environmental (spatial/temporal) variability, sampling designs can be optimized to ensure that project and data quality objectives are substantially achieved (e.g., Green, 1979; Gilbert, 1987). Alternatively, it may become apparent that the objectives are not achievable, given the available budget and existing technology, and the expense of a major program that will not "pay off" can be avoided.

Table 2 lists a few key elements that should be addressed in defining quality assurance procedures within a Quality Assurance Plan. Personnel responsible for each QA function would be identified in the Quality Management Plan. Audit activities and QA reporting would normally be assigned to a Quality Assurance Officer.

Quality control activities define the boundaries of acceptable performance for the measurement system, and include the routine checks that indicate whether the system is performing to specification. Data reporting generally stops and corrective actions are initiated when the system goes out of control. Range and average control charting methods have been described elsewhere (MOE, 1985; ASTM, 1985, 1986; Dux, 1986).

Quality assurance includes quality control, as well as a full range of in-house and external audit activities, planning and management activities, and documentation/ reporting activities that ensure accountability for data of known quality.

Data quality measures should be defined in the same terms as data quality objectives (DQO's), so that the two can be compared in program evaluation. A set of DQO's can be derived from intended data uses (e.g., hypotheses to be tested, summary statistics involved, and total uncertainty that can be tolerated). The DQO derivation process (QAMS, 1986) is usually iterative involving dialogue between data users and generators, with value judgements on both sides.

TABLE 1: SOME BASIC QA ELEMENTS

1A. SAMPLING PROCEDURES

- o Sampling Design - linked to hypotheses, DQO's
- o Preliminary Effort - permits design optimization
- o Field Equipment - maintenance, calibration, cleaning, log books
- o Collection/Preservation - field books, shipping, records, storage
- o Chain-of-Custody - signatures and dates

1B. ANALYTICAL PROCEDURES

- o Lab Facilities - receiving, storage, preparation, testing areas
- o Reagents/Solutions - labels, sources, preparation/expiry dates
- o Instruments - maintenance, calibration, log books
- o Log-In and Scheduling - link to data storage system
- o Standard Operating Procedures - general and specific methods

1C. DOCUMENTATION PROCEDURES

- o Field, Laboratory, QA/QC Methods
- o QA/QC and Routine Results
- o Access and Revision Procedures
- o Data Handling and Reporting Guidelines
  - blank, recovery correction
  - significant figures
  - T, W and other flags

TABLE 2: QUALITY ASSURANCE PROCEDURES

- **QA AUDIT ACTIVITIES**
  - review of QC sample results, performance audit results
  - review of data for logical inconsistencies, calculation/transcription errors
  - review of lab records on samples, reagents, instruments, methods
  - review of interlaboratory/accreditation performance
  - knowledge testing and training sessions

- **QA REPORTS**
  - defined schedule, signatures, long-term file

- **SAMPLING QUALITY CONTROL**
  - design sufficiency check
  - field replicates, total precision
  - trip blanks, contamination check
  - defined acceptance criteria

- **ANALYTICAL QUALITY CONTROL**
  - lab replicates, analytical precision
  - standards, accuracy assessment
  - lab blanks, contamination check
  - range and average control charts
  - out-of-control action guidelines

- **DATA QUALITY OBJECTIVES**
  - sensitivity, accuracy
  - total uncertainty, total precision, components
  - amount of valid data, representativeness

Total uncertainty tolerance (ΔC) defines an upper limit for bias plus imprecision (CI) (Taylor, 1987):

$$\Delta C \geq \text{bias} + CI$$

where CI is the confidence interval of the data summary statistic in terms of standard deviation (S) and number of samples included (n). The standard deviation may have sampling method, analytical and sometimes environmental components, which can be partitioned using field replicate and lab replicate data. Objectives can be established for each component as well as total uncertainty.

QA in Aquatic Toxicity

Toxicity measures may be considered as additional properties of the sample to be analyzed, using a test organism as the analytical instrument. Consequently, the general QA elements discussed above also apply in the toxicity laboratory (see, for example, MacGregor and Doe, 1987). However, there are unique sources of error in biological laboratories. The checklist in Table 3 includes both general and specifically biological elements. I want to focus here on some of the more important sources of error in biological testing and associated QA records and checks to be performed.

Animal sensitivity may vary with source and genetic history, age, sometimes sex, state of health, and stress associated with any recent environmental change. Diseased animals are sacrificed, and test animals are never reused. Detailed records on test animals help to determine whether they are suitable for testing, and to identify causes of any observed change in test system sensitivity. Environmental conditions such as light and temperature should be monitored for similar reasons.

Test dilution waters vary substantially from lab to lab in critical factors, such as pH and hardness, which can strongly affect test response. Water supply may also change episodically. Both water and food supplies should be routinely monitored for trace contaminants, and records maintained for comparison to acceptance criteria.

A bioassay test often involves treatment at different dose levels, and dose level selection can influence the calculated test endpoint (e.g., LC50 or EC50) as well as its associated uncertainty. Dose selection guidelines should be in place. In particular, a calculated endpoint that exceeds the dose range of the test is suspect as an extrapolation and should be flagged as such.

TABLE 3: QUALITY ASSURANCE CHECKLIST FOR AQUATIC TOXICITY
(Check or enter footnote number to describe deficiencies)

Quality Management

____ Standard Operating Procedures available (both QA and test methods).
____ Quality Management Plan for laboratory (may be part of SOP).
____ Responsibilities, authorities, qualifications defined for each position.
____ Quality Assurance Officer with adequate authority.

Quality Assurance Functions Performed

____ Review of QC sample results (permanent record available).
____ Performance assessment review (blind samples).
____ Knowledge testing and training sessions.
____ Review of test sample data (adequate checks performed).
____ Report approval mechanism (signatures).
____ Quality Assurance reports (corrective actions indicated).
____ Interlaboratory and accreditation studies (record of participation).

Laboratory Operations

____ Instrument calibration and maintenance (records, methods available).
____ Animal care and acclimation (records, methods available, tank materials).
____ Reagents and working solutions (labels, acquisition, preparation, expiry dates).
____ Glassware cleaning adequate.
____ Separation of areas (receiving, storage, preparation, testing, lab floor plan).
____ Temperature, light, ventilation control (records, adequate devices).
____ Dilution water supply (source, treatment, chemical check records).
____ Food supply (source, chemical check records).
____ Sample information system (log-in, labelling methods, holding times recorded).
____ Safety procedures adequate.

Testing, Quality Control and Reporting

____ Test record sheets (concentrations, daily responses, test conditions).
____ Dose selection guidelines (range tests, concentration intervals).
____ Dose verification (frequency, methods for mixtures).
____ Dilution water controls (use of data, test acceptance criteria).
____ Test replication (frequency, range control limits).
____ Reference toxicant tests (type, frequency, control limits).
____ Control charts or tables (limit determination, out-of-control actions).
____ LC50 and other endpoint calculations (method selection guidelines).
____ Data reporting (confidence limits, methods, test conditions, QC data).

Deficiency Descriptions

Footnote
No. ____ Problem

Chemical dose verification during a test is always desirable, preferably at the beginning and end. However, it adds to test cost and is often not routine. Moreover, if the test agent is an uncharacterized mixture, a surrogate such as conductivity or dissolved organic carbon must be used for dose verification. Verification frequencies and methods for mixtures should be defined.

Dilution water (negative) controls should be part of every test, and a zero response is expected. An upper limit of acceptable control response should be defined. The issue of control (i.e., blank) correction will be familiar to the chemist. In toxicity, the correction involves Abbott's formula (Finney, 1981) and may or may not be applied. It probably makes little difference when control response is low, but the control response and any control correction should always be stated with test results.

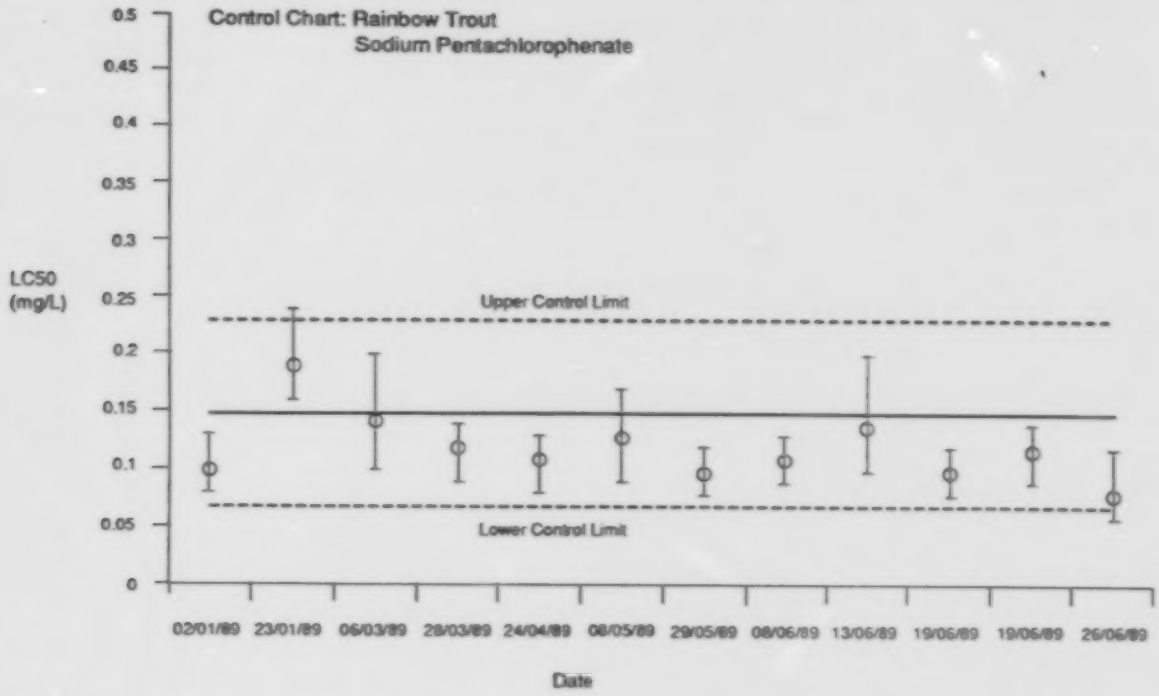
Reference toxicant tests are positive controls, and should be performed frequently. Recent Environment Canada protocols specify monthly tests (Environment Canada, 1989a,b,c). Reference toxicants are roughly equivalent to the internal calibration or control standards used by the chemist. Unfortunately, there is nothing resembling a Certified Reference Toxicant. Either reference toxicant tests or split sample exchanges can be used to address comparability between laboratories (e.g., Grothe and Kimerle, 1985; PACE, 1989), but not to decide which lab is correct.

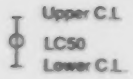
Various methods of LC50 or other endpoint calculation are in use (Stephan, 1977), and they do not always give the same results. Some have more stringent data requirements and cannot always be used. Method selection guidelines should be in place, and methods should be stated with test results.

Both average and range type control charts can be used in control of bioassay measurement systems (EPA, 1978). Reference toxicant data can be used on both, if replicate reference tests are performed. Alternatively, routine samples can occasionally be split for replicate testing. An uncertainty measure (confidence limit) is obtained from the dose response curve of any single test, even without test replication, however range charting provides a check on repeatability, and a visual indicator of developing test precision problems.

Figure 1 illustrates the use of average control charts for reference toxicant test data. Figure 2 illustrates several approaches to range charting of replicate test data, over the full range of test response. When the precision function is non-linear or biphasic, multiple charts can be used, each controlling a narrower response range. Separate charts can also be used for different sample types.

FIGURE 1
Average Control Charts for Reference Toxicants





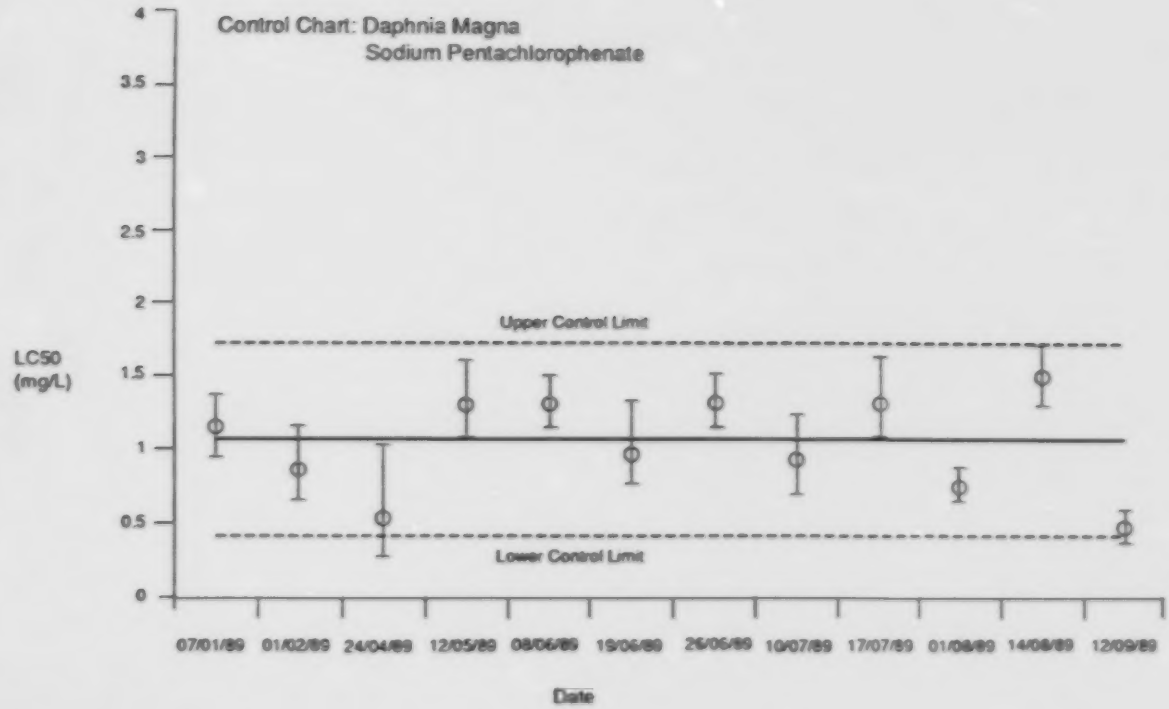
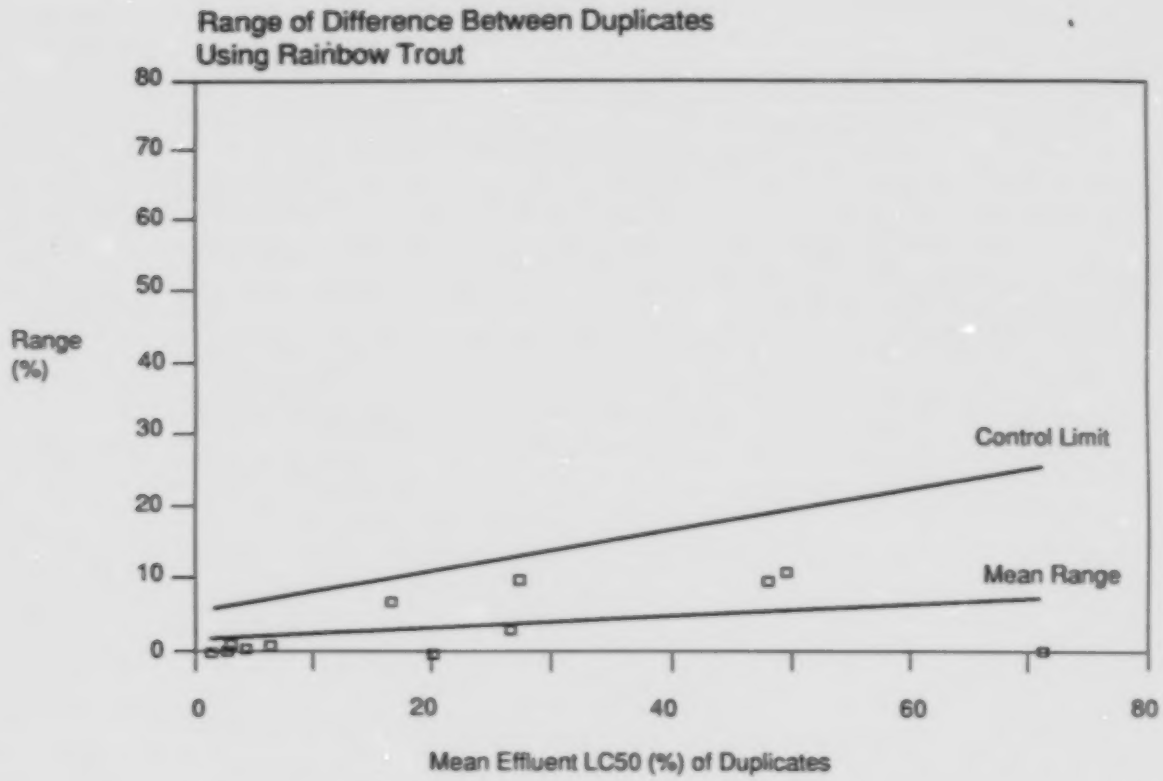
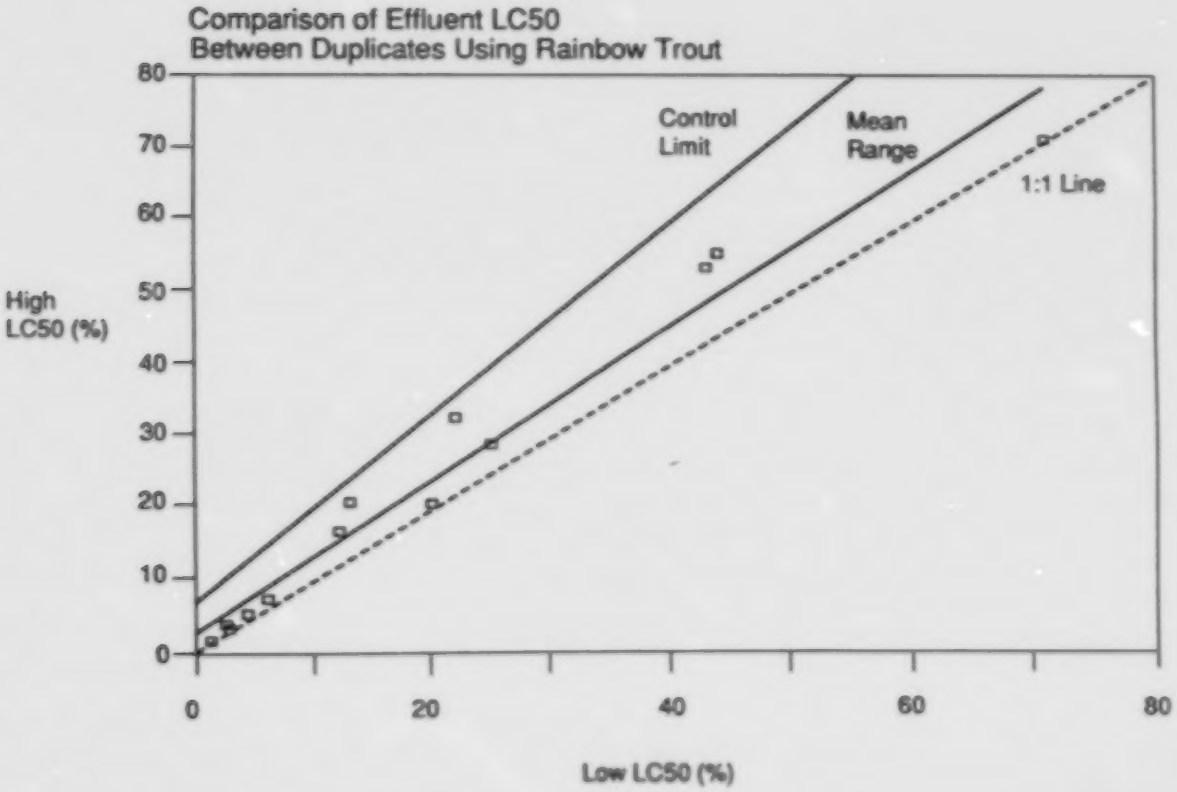


FIGURE 2
Control Charting of Duplicate Toxicity Test Data





QA In Aquatic Field Studies

A complete biological assessment of pollution impacts on the aquatic environment involves more than toxicity testing of effluent samples. In situ toxicity tests in the receiving water are more definitive for evaluation of short-term effects, while ecological assessment of benthic, plankton and fish communities are more definitive for evaluation of long-term consequences. Benthic communities, because of their sedentary nature, are particularly useful in revealing spatial patterns of biological effect related to point sources of pollution.

A wide variety of sampling and analytical methods are used in ecological community assessments, according to the type of organism considered, the specific study objectives, and the personal preference of the investigator. With each approach, there are different specific sources of error to be minimized, if possible, and quantified for purposes of data quality evaluation. However, there are common elements, and therefore common general approaches to quality assurance. I will attempt to focus on these. For example, most ecological assessments involve sampling, taxonomic identification and enumeration of organisms. We may be using these data to estimate species biomass, abundance, relative abundance, diversity and/or trophic structure of the community.

The checklist in Table 4 includes the same general QA elements that were previously mentioned, and listed at the top of the aquatic toxicity checklist. The principles are no less important here, although formal interlaboratory and accreditation studies may be rare. In addition, I have added some new concepts of particular relevance in field ecological assessment, and will discuss these briefly.

Development of an optimum sampling design requires preliminary data, to determine, for example, what is an adequate level of sampling effort at each sampling point in time or space, and how many sampling points are needed. The return of information increases with increasing effort, up to a point of limited return, and this point can be defined (Smith, 1974; Stevenson and Lowe, 1983). With inadequate effort we may consistently underestimate the number of species present, or achieve inadequate precision on density estimates. Distribution of the preliminary sampling in space will help to identify any environmental heterogeneities present (i.e., strata), which will then help to optimize design. Sampling and environmental variance can be estimated, and used to pre-determine the appropriate number of samples for the study (or for each stratum), based on project and data quality objectives (e.g., Elliot, 1971). In some cases, when sample processing can be completed rapidly in the field, (e.g., a fish population estimate, Peck et al., 1988), a sequential sampling approach can be used, whereby sampling continues until precision objectives are known to be achieved.

TABLE 4: QUALITY ASSURANCE CHECKLIST FOR AQUATIC FIELD STUDIES
(Check or enter footnote number to describe deficiencies)

Quality Management

___ Standard Operating Procedures available (both QA and test methods).
___ Quality Management Plan for institution (may be part of SOP).
___ Responsibilities, authorities, qualifications defined for each position.
___ Quality Assurance Officer with adequate authority.

Quality Assurance Functions Performed

___ Review of QC sample results (permanent record available).
___ Performance assessment review (blind samples).
___ Knowledge testing and training sessions.
___ Review of test sample data (adequate checks performed).
___ Report approval mechanism (signatures).
___ Quality Assurance reports (corrective actions indicated).
___ Interlaboratory and accreditation studies (record of participation).

Field Operations

___ Preliminary sampling (return for effort, environmental strata, variability)
___ Sampling design (based on objectives, adequate replication)
___ Instrument calibration and maintenance (records, methods available)
___ Crew training and evaluation (video records and comparisons)
___ Sampling equipment (appropriate, consistent, cleaning and maintenance records)
___ Collection, preservation and storage (methods, adequate labels, records)
___ Additional effort checks (confirm little gain of information)

Sample Processing and Reporting

___ Sorting and subsampling (methods, adequate records, error estimates)
___ Sorting verification (confirm few organisms overlooked)
___ Taxonomy and enumeration (appropriate keys, records, reference collection)
___ Duplicate processing (within and between personnel, error estimates)
___ Exchange samples (between laboratories, comparability assessed)
___ Taxonomic verification (by recognized experts)
___ Archives (for samples, voucher specimens)
___ Data handling and reporting (entry checks, missing value codes)

Deficiency Descriptions

Footnote
No. Problem

Crew training and evaluation are particularly important in field studies where written sampling protocols, if they exist, are often lacking in detail and open to some interpretation (Adams et al., 1983). Video recordings can be a useful tool for standardization of field technique, although they are no substitute for written methods. When several contractors are involved, videos can be required to check and document consistency of technique.

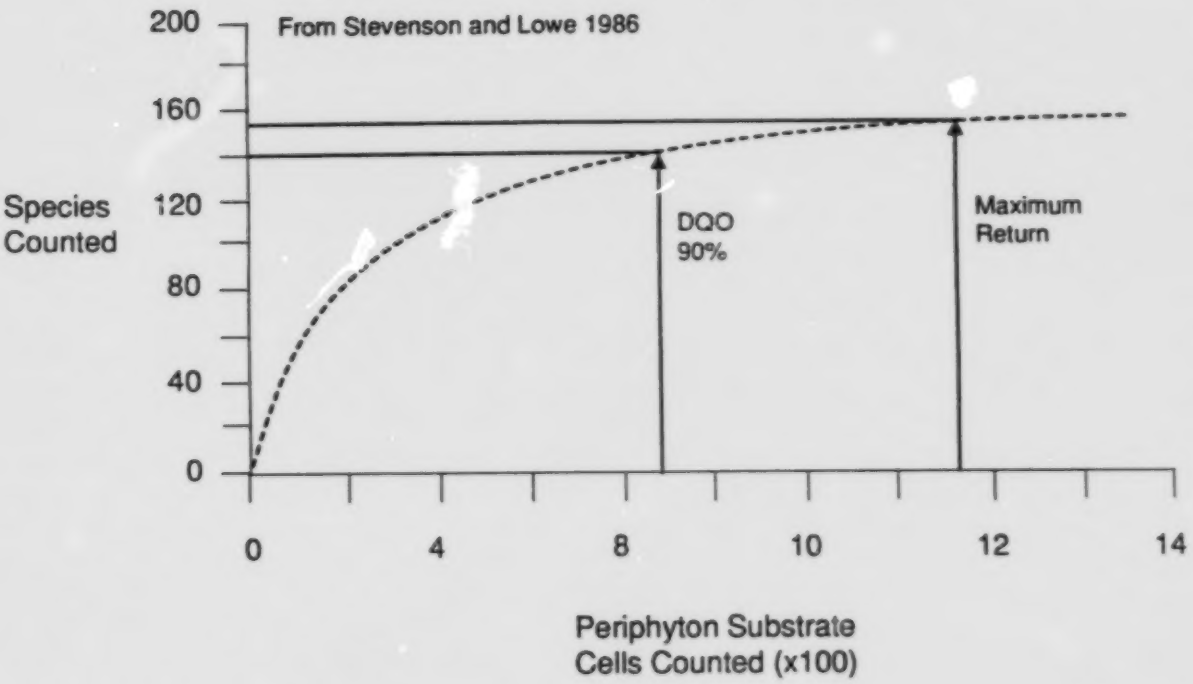
Methods will specify sampling equipment appropriate to study objectives, and call for consistency within a project. Consistency in equipment is important since different sampling devices have different biases in terms of collection efficiency and type of organism captured. Methods should also specify proper use of equipment, location of stations, labelling and preservation of samples, and any constraints on sample storage.

Additional effort checks are performed in the field, during the routine survey, and serve to confirm the return-for-effort anticipated on the basis of preliminary field work. For example, if a data quality objective is to retrieve 95% of benthic species present at any station, and it has been estimated then 3 Ponar grabs per station are sufficient to achieve this, more grabs can be collected at a few selected stations until the number of species reaches a plateau, and it can be determined whether the first three were sufficient to achieve the 95% objective. The same approach can be applied when sampling plankton (return vs. number of plankton tows) or fish (return vs. number of seine net passes).

Sample processing may involve, as a first step, sorting of organisms from debris, and possibly subsampling of sorted organisms for detailed identification. As with sample collection, written methods for sample processing are required, records should be kept, errors (i.e., variance) associated with these activities should be estimated (e.g., Kreis, 1986, 1989). Such error estimation necessarily involves some type of duplicate processing, either on split samples, or by re-mixing and reprocessing the same sample (Peck et al., 1988). There are problems associated with each approach (e.g., obtaining an even split or a thorough mix) but these should not preclude at least some effort to characterize processing error, both within and between personnel. Samples can also be exchanged between laboratories for duplicate processing when several labs are involved in a study.

Verification of sorting efficiency is easily performed on a spot check basis if the left-over debris from a sample is retained. Stains can be applied to make any organisms that were missed stand out from the debris. This is particularly useful with benthic samples, which usually contain large quantities of debris. Contract laboratories are much more likely to sort thoroughly when they know that samples will be randomly checked for sorting efficiency.

FIGURE 3
Designing Effort to a Species Inventory Objective



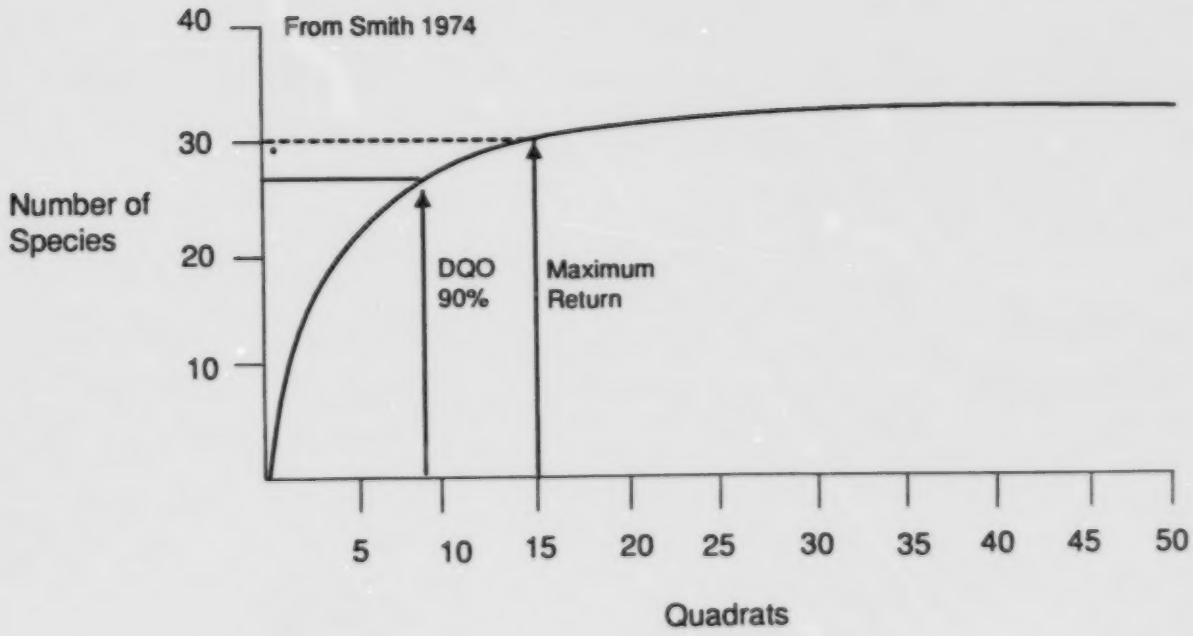
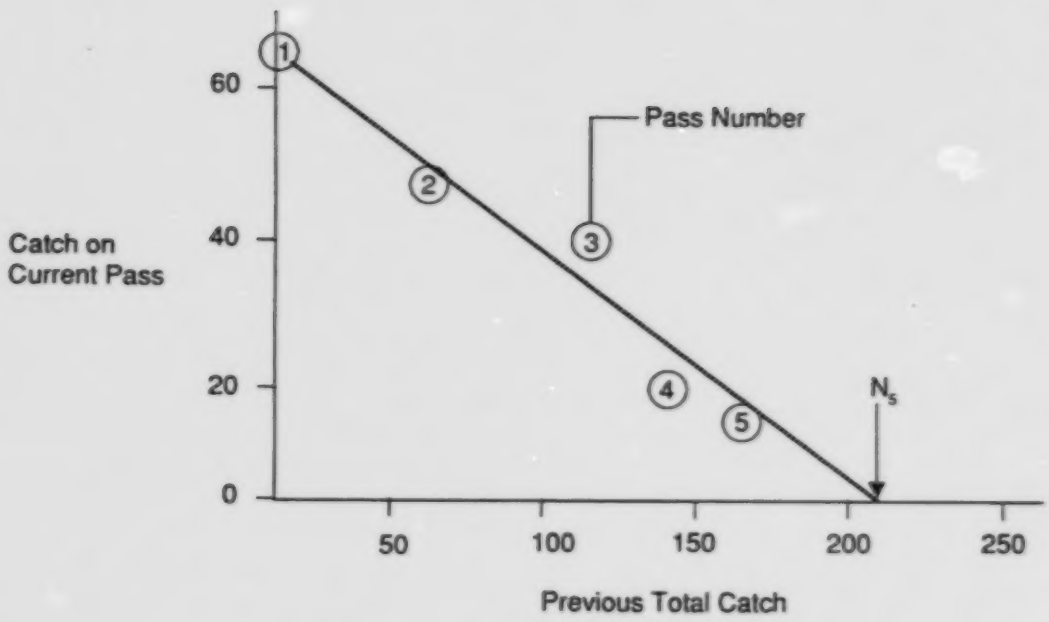
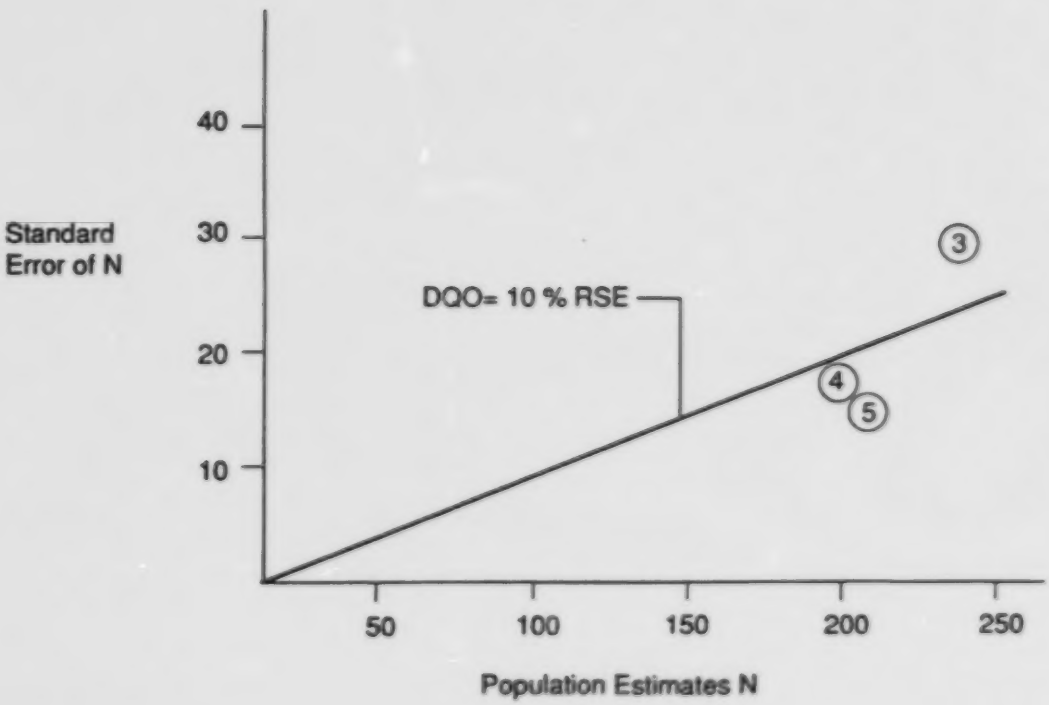


FIGURE 4
Sequential Sampling to Meet Precision Objective





Taxonomic methods will refer to recent published keys, and also describe staining and mounting techniques prior to identification. Keys must be appropriate to the geographic region of study, and should always be cited in data reports. Computerized keys have been developed. Reference collections of organisms should be maintained by each taxonomic lab, with master collections when studies involve several regional laboratories. Photographic iconographs have been used to advantage (Camburn et al., 1984). These collections are useful in confirming identifications and ensuring consistent taxonomy, as well as training personnel.

Verification of taxonomy by recognized experts should be performed on a spot check basis, and on any new specimen entering a reference collection. Museums are sometimes prepared to perform this service when remote areas are included in the study, and new species or distribution records are likely. They will also sometimes serve as archives for selected samples and/or voucher specimens of organisms. However, since considerable effort is involved in long-term maintenance of preserved biological material, it is often necessary to minimize the quantity of material submitted.

Ecological community studies usually generate large data matrices, since hundreds of species may be involved, and an observation on each is recorded at every sampling station. Computers and multivariate statistical programs generally assist in synthesis of the data to elucidate spatial and/or temporal patterns of community health (e.g., Hensler, 1982; Ludwig and Reynolds, 1988). Due to the nature and quantity of data, particular care is needed in data handling, to guard against data entry errors or subsequent misuse of data. Double entry systems and transcription checks against the original data records are useful QC techniques. Missing data must be clearly distinguished from a species absence by use of unique non-zero missing value codes, with code definitions built into each file. Read-only files help to ensure data integrity.

Figure 3 illustrates the return-for-effort concept, as utilized in aquatic study design to achieve an adequate species inventory. Figure 4 illustrates a case where sequential sampling can be used to achieve adequate precision on a population estimate.

Summary

Regulatory use of biological data will require increased QA/QC effort to ensure data of known quality, and to permit demonstration of adequate data quality. The basic principles of quality assurance and quality control are the same in biological and chemical measurement systems. However, the applications to biological measurement are not yet standardized, particularly for field activities. Specific applications will undoubtedly evolve over the next few years. We hope that our guideline document will encourage some consistency in approach, without impeding development of new ideas.

References

ASTM. 1985. Standards on Precision and Bias for Various Applications. Second Edition. American Society for Testing and Materials, Philadelphia.

ASTM. 1986. Manual on Presentation of Data and Control Chart Analysis. Committee E-11 on Statistical Methods. ASTM Special Technical Publication 15D. American Society for Testing and Materials, Philadelphia.

ASTM. 1988. Draft Standard Practice for Generation of Environmental Data. Committee E-47 on Biological Effects and Environmental Fate. Subcommittee E47.05 on Quality Assurance. American Society for Testing and Materials, Philadelphia.

Adams, W.J., T.J. Hoogkeem and P.R. Michael. 1986. Aquatic Monitoring: A Rationale for Obtaining and Interpreting Aquatic Ecosystem Chemical Exposure Data, In: Isom, B.G. (Ed.), Rationale for Sampling and Interpertation of Ecological Data in the Assessment of Freshwater Ecosystems, STP 894. American Society for Testing and Materials, Philadelphia.

Camburn, K.E., J.C. Kingston, and D.F. Charles (eds.). 1984-1986. PIRLA Diatom Iconograph. Contains 53 photographic plates and 1059 figures, plus figure legends. PIRLA Unpublished Report Series, Report number 3. Indiana University, Bloomington, IN.

Day, K.E., E.D. Ongley, R.P. Scroggins and H.R. Eisenhauer. 1988. Biology in the New Regulatory Framework for Aquatic Protection. Proceedings of the Alliston Workshop, Environment Canada, Ottawa.

Dux, J.P. 1986. Handbook of Quality Assurance for the Analytical Chemistry Laboratory. Van Nostrand Reinhold Co., New York.

EPA. 1978. Quality Assurance Guidelines for Biological Testing. Prepared for EPA Environmental Monitoring and Support Lab, Las Vegas, by Tractor Jitco, Inc., Rockville, Maryland. PB-285 369.

Elliot, J.M. 1971. Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates. Scientific Publication No. 25, Freshwater Biological Association, Westmoreland, U.K.

Environment Canada. 1989a. Acute Lethality Test Using Rainbow Trout (Salmo gairdneri). Final draft, April 1989. Environmental Protection Service, Ottawa.

Environment Canada. 1989b. Acute Lethality Test Using Daphnia sp. Final draft, April 1989. Environmental Protection Service, Ottawa.

Environment Canada. 1989c. Acute Lethality Test Using Threespine Stickleback (Gasterosteus aculeatus). Final draft, April 1989. Environmental Protection Service, Ottawa.

Finney, D.J. 1981. Probit Analysis. Cambridge University Press, Cambridge.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Co., New York.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley and Sons, New York.

Grothe, D.R. and R.A. Kimerle. 1985. Inter- and intralaboratory variability in Daphnia magna effluent toxicity test results. Environmental Toxicology and Chemistry 4: 189-192.

Hensler, G.L. 1984. An Introduction to Statistical and Ecological Software, In Gertz, S.M. and M.D. London (Eds.), Statistics in the Environmental Sciences, STP 845, American Society for Testing and Materials, Philadelphia.

IJC. 1988. Quality Management Plan for Great Lakes Internationial Surveillance with Implementation Guidelines. May 1988. International Joint Commission (draft).

Kreis, R.G., Jr. 1989. Variability Study - Interim Results. Section 4 In: Charles, D.F. and D.R. Whitehead (Eds.). Paleoecological Investigation of Recent Lake Acidification (PIRLA): 1983-1985: Interim Report, October 1989. Electric Power Research Institute, Palo Alto, CA.

Kreis, R.G., Jr. 1986. Variability Study. Section 17 In: Charles, D.F. and D.R. Whitehead (Eds). Paleoecological Investigation of Recent Lake Acidification (PIRLA): Methods and project description. Electric Power Research Institute, Palo Alto, CA.

LRTAP. 1986. Long-Range Transport of Air Pollutants. Quality Management Plan. Federal/Provincial Research and Monitoring Coordinating Committee.

Ludwig, J.A. and J.F. Reynolds. 1988. Statistical Ecology: A Primer on Methods and Computing. John Wiley and Sons, New York.

MOE. 1985. Laboratory Services Branch, Quality Management Plan. Ontario Ministry of the Environment, Rexdale.

MOE. 1984. Principles of Control Charting. D.E. King and G.C. Ronan, Laboratory Services Branch, Data Quality Report Series. Ontario Ministry of the Environment, Rexdale.

MacGregor, D. and K. Doe. 1987. Basic Elements of Quality Assurance in Ecotoxicological Testing. In: G.A. Sergy (Ed.). Recommendations on Aquatic Biological Tests and Procedures for Environmental Protection, C&P, DOE. Conservation and Protection, Environment Canada, Ottawa.

OECD. 1981. Decision of the Council Concerning Mutual Acceptance of Data in the Assessment of Chemicals. Annex 2. OECD Principles of Good Laboratory Practice. In: Guidelines for Testing of Chemicals. Organization for Economic Cooperation and Development, Paris.

OECD. 1989. Draft Council Decision-Recommendation on Compliance with Principles of Good Laboratory Practice (restricted). Organization for Economic Cooperation and Development, Paris

QAMS. 1980. Interim Guidelines and Specifications for Preparing Quality Assurance Project Plans. Quality Assurance Management Staff, Environmental Protection Agency, Washington. EPA-600/4-83-084. PB-83-170514.

QAMS. 1986. Development of Data Quality Objectives. Description of Stages I and II. Quality Assurance Management Staff, Environmental Protection Agency, Washington (draft).

Smith, R.L. 1974. Ecology and Field Biology. Harper and Row, New York.

Stephan, C.E. 1977. Methods for calculating an LC50. In: Aquatic Toxicology and Hazard Evaluation. ASTM STP 634. (F.L. Mayer and J.L. Hanelink, Eds.). American Society for Testing and Materials, Philadelphia.

Stevenson, R.J. and R.L. Lowe. 1986. Sampling and Interpretation of Algal Patterns for Water Quality Assessments. In: Isom, B.G. (Ed.), Rationale for Sampling and Interpretation of Ecological Data in the Assessment of Freshwater Ecosystems, STP 894. American Society for Testing and Materials, Philadelphia.

Taylor, J.K. 1987. Quality Assurance of Chemical Measurements. Lewis Publishers, Inc., Chelsea, Michigan.

BIOLOGICAL QUALITY ASSURANCE FOR ECOLOGICAL MONITORING: CAN WE USE CHEMISTRY AS A TEMPLATE?

James E. Pollard and David V. Peck
Lockheed Engineering and Sciences Company
1050 East Flamingo Road
Las Vegas, Nevada 89119

Notice

Although the information in this paper has been funded wholly or in part by the U.S. Environmental Protection Agency under Contract Number 68-03-3249 to Lockheed Engineering and Sciences Company, it has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred. The mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Abstract

The conceptual basis for inclusion of various types of quality assurance/quality control (QA\QC) material during sample processing has been well documented for environmental chemical analyses (Taylor, 1987). The U. S. Environmental Protection Agency (EPA) has defined five aspects of data quality that must be addressed in any data collection effort funded by the EPA: precision, accuracy, completeness, comparability, and representativeness. Although these conceptual aspects of chemical data quality are generally definable using various types of QA\QC samples, biological quality control and documentation procedures are not well defined at present. Within this paper, the conceptual basis of control and documentation points during sample collection and processing will be presented using analytical chemical analyses from the National Surface Water Survey (NSWS) and a follow-on acid deposition study, the Episodic Response Project (ERP), as examples. A hierarchical variance component model is presented using the QA sample design and results from the ERP. Control points for one type of biological sample (benthic macroinvertebrates) are presented and discussed along with a proposed variance component model analogous to that presented for the ERP. Other types of field-based biological sampling procedures are discussed and potential QA\QC for these activities explored. Finally, the potential evolution of biological QA is examined with some suggestions for future research.

Introduction

The conceptual basis for inclusion of various types of quality assurance/quality control (QA\QC) samples during sample processing has been well documented for analytical chemical analyses (Taylor 1987). The U. S. Environmental Protection Agency (EPA) has defined five indicators of data quality that must be addressed in any data collection effort funded by the EPA: precision, accuracy, representativeness, comparability, and completeness (Stanley and Verner, 1983). Although these conceptual aspects of chemical data quality are generally definable using various types of QA\QC samples, quality control and documentation procedures for biological data collected in ecological monitoring studies are not well defined. It is well known that control material can be used to maintain statistical control of the measurement process in an analytical laboratory. The analogous procedures for biological sample processing are not as obvious. For example, what is the analogue to a standard reference sample in a field fisheries study? We can take standards to the field to calibrate field chemical measurements (e.g. pH buffers), but in many cases we cannot envision a comparable process to ensure the quality of field biological measurements. Within this paper we will provide examples of typical control points used to assure data quality in environmental chemistry surveys, and will describe our initial attempts to carry these concepts into the biological realm.

Control Point Concepts

The concept of control and documentation points in sample processing during environmental surveys can be illustrated using the Aquatic Effects Research Program (AERP) as an example. This program was implemented in 1983 by the EPA's Office of Research and Development to provide an integrated, stepwise approach to resolution of issues related to the concern over the effects of acidic deposition on U.S. surface waters. A number of component research projects were implemented under the AERP between 1983 and the present, one of which was the National Surface Water Survey (NSWS). This effort was a large scale chemical survey of the status and extent of acidification of the Nation's surface waters. Control points in the measurement process documented during the NSWS were: (1) field sampling; (2) field processing (aliquoting and preservation); (3) shipping and handling; and (4) laboratory analyses, both within and between laboratories.

A unique aspect of the NSWS was the effort that was expended on the quality assurance (QA) program for the surveys. Stringent quality control (QC) protocols were developed and written into the contracts granted to analytical laboratories. Control of the measurement process was maintained and documented internally using classical QC material such as calibration standards, quality control check standards, laboratory blanks, and duplicate analyses. Because of the scale of the surveys, it was necessary to contract with multiple laboratories in order to process the sample loads generated within the specified time constraints. Therefore, it was inevitable that a certain degree of interlaboratory bias, as well as other sources of error, were introduced into the data

as a result of the various analytical and processing activities. This error was documented by analysis of an extensive series of double blind QA samples (completely unknown to the analyst as performance samples) that were added to groups of samples (batches) submitted to the analytical laboratories.

Double blind QA samples used in the NSWS were field blanks, field duplicates, field audits, and laboratory audits. Field blank samples were designed to document field contamination problems and to provide a mechanism to determine the system background levels. Field duplicate samples were designed to determine overall within-batch variability, including field handling and processing. Field audits were designed to document among-batch variability, including handling and processing at the field processing laboratory, where the samples were batched. Laboratory audits were designed to document the among-batch error introduced by laboratory analytical procedures only. The general conclusions reached during the NSWS were that field sampling, field contamination, and within-batch error were minor elements of the sample processing error (Pollard and Heggem, 1989). Among-batch error within a single laboratory represented the largest component of the measurement error process, while some evidence existed that interlaboratory bias could be a significant problem with some analytes (Mitchell-Hall et al., 1989).

The Episodic Response Project (ERP) is a program within the AERP designed to assess the importance of short-term, event-based acidification to surface water chemistry, hydrology, and aquatic biota. Based on the NSWS experience, among-batch error (differences between calibration runs) was considered a primary control point, as was interlaboratory bias. A major portion of the QA program for the ERP centers upon internal quality control performed on-site by each cooperating agency. In addition, the program emphasis focuses on extensive use of natural matrix audit material that is well characterized for both control and documentation samples.

Control Points and Variance Components

The concept of control points within a sample processing stream can be illustrated by viewing the results of natural audit sample analyses performed within the ERP using a hierarchical experimental design. Well characterized natural audit material was introduced at various stages of sample processing to determine the relative magnitude of variance components associated with sample processing steps (Figure 1). Separate aliquots of the natural audit material were sent to the five laboratories participating in the ERP. One of the laboratories sent an aliquot of the audit to the field to be processed through the sampler and returned to the laboratory to be processed within the same analytical batch as a second aliquot which was held in the laboratory under controlled conditions. Both field treated and laboratory held aliquots were analyzed in duplicate within a single batch. In addition, the test laboratory held an aliquot of the audit for analysis in a later batch of samples. The other laboratories processed one aliquot of the audit during the initial time period.

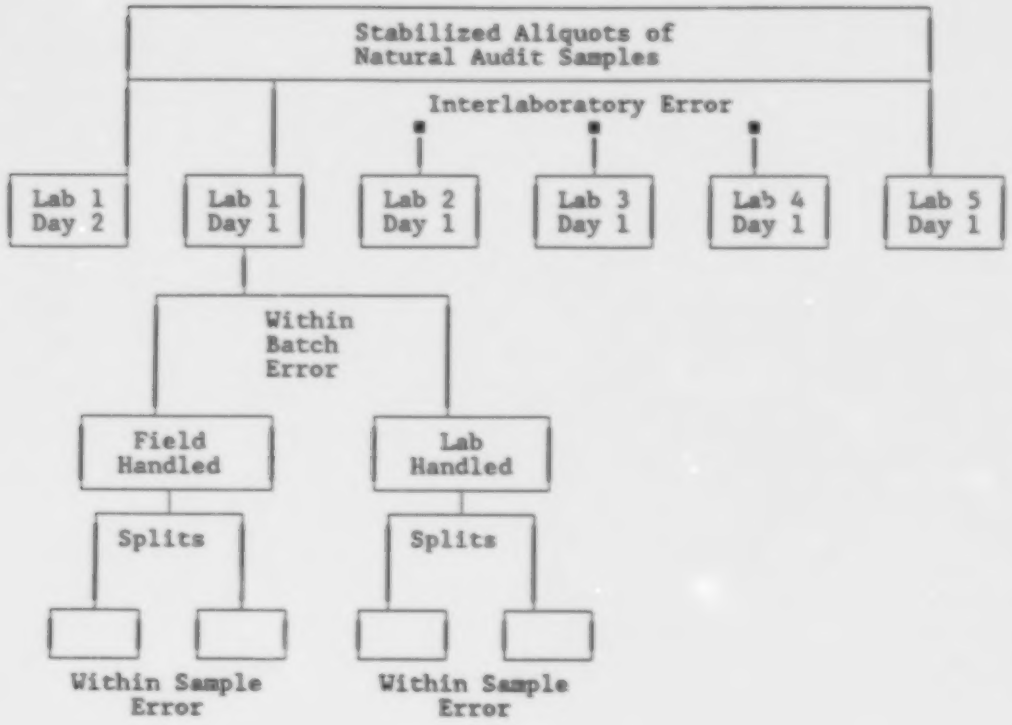


Figure 1. Variance component design illustration using natural audit material.

Results of these tests were variable, depending on the analyte of interest. Two examples illustrating the types of information that can be gained by this type of analysis can be found in Table 1. With calcium, the major variance component was between laboratories with among-batch variability being secondarily important. Therefore, if obtaining a better estimate of calcium concentrations was an important program goal, then more effort would need to be assigned to controlling interlaboratory bias. On the other hand, interlaboratory bias for ANC was negligible, while among-batch bias within a laboratory was the major error component indicating calibration as the major control point. Within this program, predefined data quality objectives (DQOs) were 10 percent total error for calcium, and ± 10 μeq/L for ANC. Both of these objectives were met in the examples presented, indicating that no corrective action needed to be taken. However, these examples do illustrate the concept of defining control points and identification of the most important ones within a sample processing stream.

Table 1. Examples of Relative Percent Variance Components for Calcium and Acid Neutralizing Capacity

Variance Component	% For Ca[a]	% For ANC[b]
Interlaboratory	87	0
Among Batch	13	83
Within Batch	0	7
Within Sample	0	10
Means and Confidence intervals		
Mean Concentration	1.8 mg/l	(-) 1.0 μeq/l
Confidence Interval for Total Error	± 0.19mg/l	± 6.6 μeq/l

a - Ca = Calcium
b - ANC = Acid Neutralizing Capacity

Control Points in Biological Sample Processing

For many types of biological samples, it is difficult to envision what a control point in sample collection and processing might be. Macroinvertebrate samples are one type of biological sample that provides a classical example of a well known sample type that has had little documented quality control associated with it. While standard operating procedures and guidance for quality assurance of biological testing has been available for quite some time (U.S. EPA 1978), only recently has any quality control or documentation been specified in the literature (Pollard and Melancon, 1984; Chapman et al., 1987; Peck et al., 1988; Plafkin et al., 1989). There are numerous steps in the collection and processing of these samples where quality control checks can be applied or used to document data quality (Figure 2).

The importance of maintaining control of these steps can be illustrated using data from a methods research experiment funded by the EPA in 1982 (La Point et al., 1983). During this study, five replicate, standardized benthic grab samples were collected along a metal pollution gradient ranging from unpolluted to heavily polluted. Sample processing during this study utilized strict quality control for most of the steps presented in Figure 2 including field sample integrity checks, sorting quality control checks, and data entry completeness checks. However, no verification by a knowledgeable biologist was performed on the final electronic data base.

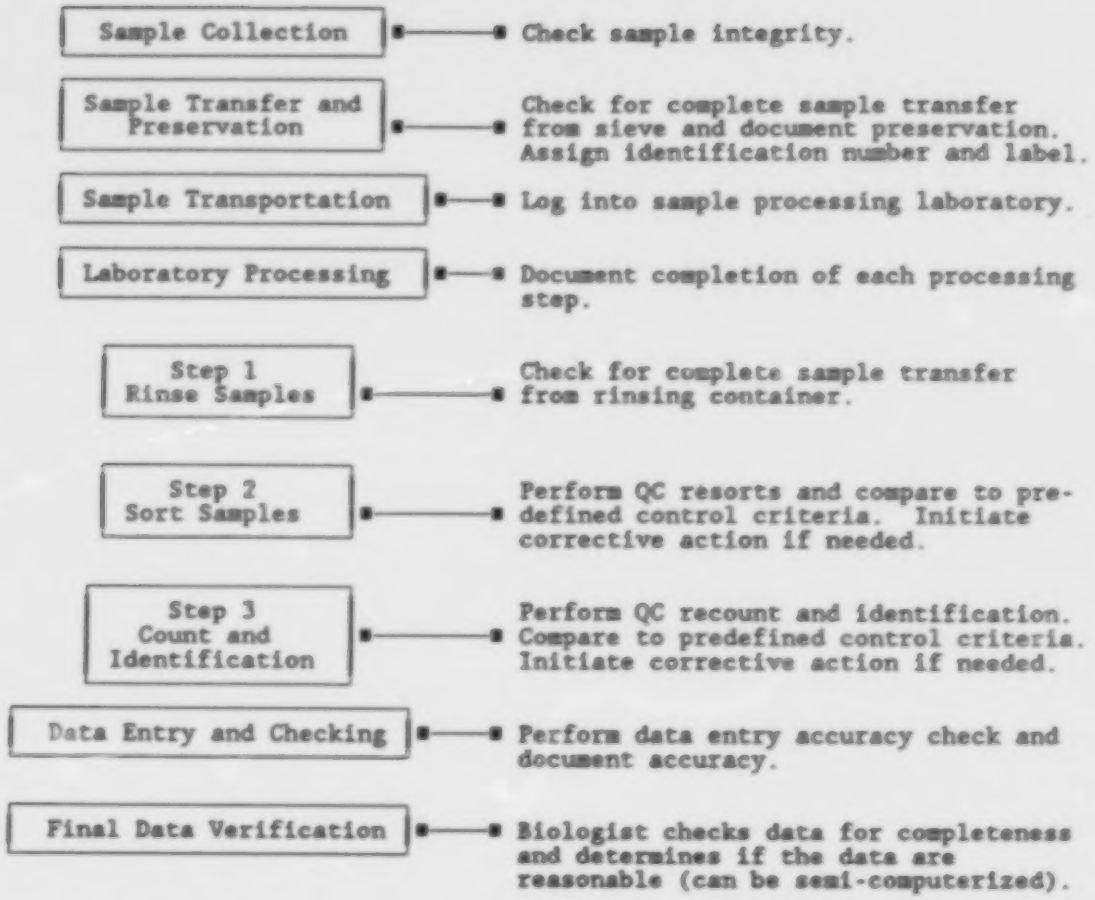

Sample Collection
Check sample integrity.
Sample Transfer and Preservation
Check for complete sample transfer from sieve and document preservation. Assign identification number and label.
Sample Transportation
Log into sample processing laboratory.
Laboratory Processing
Document completion of each processing step.
Step 1 Rinse Samples
Check for complete sample transfer from rinsing container.
Step 2 Sort Samples
Perform QC resorts and compare to predefined control criteria. Initiate corrective action if needed.
Step 3 Count and Identification
Perform QC recount and identification. Compare to predefined control criteria. Initiate corrective action if needed.
Data Entry and Checking
Perform data entry accuracy check and document accuracy.
Final Data Verification
Biologist checks data for completeness and determines if the data are reasonable (can be semi-computerized).

Figure 2. Benthic macroinvertebrate sample quality control check points and QA documentation points.

Counts for total number of organisms for the five replicates at the unpolluted site indicated one sample with very low counts. Since small area benthic grab samples (0.1 m^2) often exhibit high variance in total counts, a scan of total counts would not alert the data entry technician to any potential error in the data. Later analysis of counts for general groups of organisms (Table 2) indicated that the sample with low counts was missing entire groups rather than being generally low in counts. Since the groups that are missing (Ephemeroptera, Plecoptera, and Chironomidae) would have been sorted into separate vials, it is highly likely that these vials were misplaced and not counted. If a final verification had been performed by a biologist or computer software had been developed to make checks for unreasonable data (e.g. checks for missing major groups), this error could have been corrected at that time. The consequences of not including a final verification check in the sample processing stream resulted in very high replicate variance for this set of samples. This produced non-representative estimates of

community composition at this site resulting in highly reduced statistical power for determining differences between stations.

Table 2: Summary of Data Presented for Station 1 Prickly Pear Creek, Montana (Extracted from La Point et al. 1982)

Taxonomic Group	Rep1	Rep2	Rep3	Rep4	Rep5
Ephemeroptera	397	243	276	456	0
Plecoptera	11	8	20	14	0
Trichoptera	85	10	22	17	55
Chironomidae	97	36	29	60	0
Other Diptera	23	30	11	39	5
Coleoptera	4	32	23	20	5
Oligochaeta	0	0	19	10	0
Total	617	359	400	616	65

Variance Components in Biological Sample Processing

While we can look back at what was, for the time, a state-of-the-art benthic research laboratory and find fault in our quality control procedures, it should be kept in mind that only recently has QA\QC in biology been specified in research plans. In fact, it can be somewhat disturbing to think about being able to identify a QC error in data which had an extensive QA check system in operation when most data more than two to three years old have no mention of any QA/QC associated with collection, processing, or data base management. We would not like to speculate about the percentage of biological programs today that have operational QA/QC check systems built into them.

We have demonstrated the potential in chemical analysis of identifying significant control points using a variance components approach. Biological sample processing can have very close analogues to chemical sample processing variance components. Again we can use the benthic sample collection and processing as an example. A performance audit could be created by collecting a large benthic sample from a fairly rich habitat. This sample, split in the laboratory using documented splitting techniques, creates an analogue to a natural water sample audit. This sample could then be used in an analogous way such as that depicted in Figure 2. Prior to an environmental evaluation program the major control points could be identified for the sorting process as well as the counting and identification process (Figure 3). Control points which could adversely affect use of data for predefined objectives could be identified and corrective measures

employed to maintain sample processing control. In addition, this type of sample could be used occasionally during the program to document overall error in the measurement process.

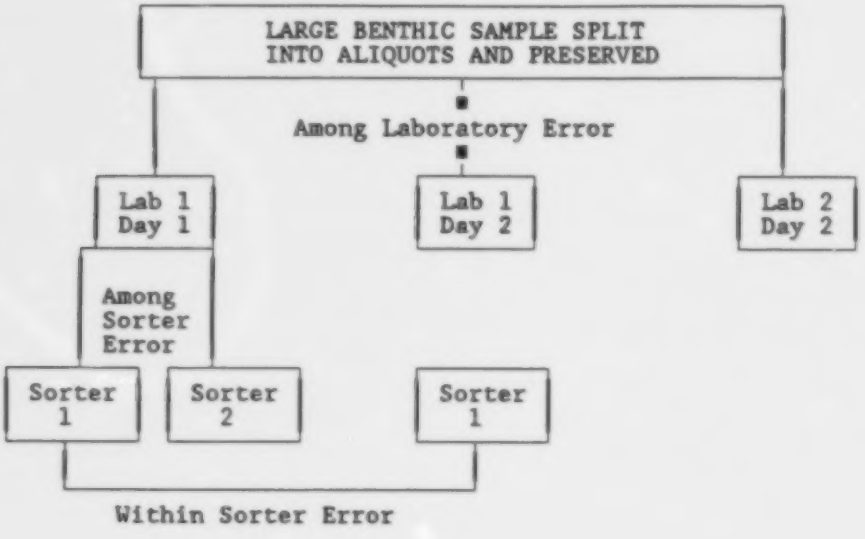


Figure 3. Variance component design for sorting of benthic samples. Analogous processes could be implemented for identification and counting.

Where Do We Go From Here?

We have presented a fairly simple case for field-based biological samples and how we can initiate a quality control and documentation program for a specific sample type. The design of effective biological QA/QC programs for data that are not generated in a laboratory is much less obvious. Below, we present a partial list of chemical QA/QC sample types that have been used to control and document data quality for environmental surveys (Table 3). Some of these samples have analogous biological counterparts, and some do not. In addition to the biological counterparts presented in Table 3, there are a number of biological measurements that are taken frequently in the field for which QA/QC procedures will need to be developed in the future. Some of the more difficult ones that we would like to present as food for thought are: population estimate data (e.g. fish, crawfish, birds, mammals, plants); length/weight data (e.g. fish, larval insects); and in-situ bioassay data (e.g. caged animals in potentially toxic environments).

Table 3. Types of Quality Control/Quality Assurance Samples Used in Analytical Chemistry Studies and Biological Analogues

Chemical QA/QC Samples	Macroinvertebrate QA/QC Samples
Laboratory Blank/ Field Blank	This type of sample would serve little purpose for evaluation of contamination potential.
Laboratory Splits	Laboratory generated subsamples.
Field Splits	Field generated subsamples.
Field Replicate	Field replicate.
Laboratory Spike/ Field Spike	Synthetically prepared samples with known and variable numbers of organisms added.
Calibration Standard	Reference collections.
QC Check Standard	Blind Pre-Identified sample resubmitted.
Performance evaluation Sample	Large sample split and stored for inclusion in the sample stream as blinds.

Conclusions

In conclusion, we feel that many local or small scale ecological monitoring studies are "quality assured" to some extent, at least in terms of study design and documentation of some level of variability. However, this documentation is not always published with the study results, or is not specifically termed "quality assurance" or "quality control" information, particularly when related to biological data. As we move into heavier reliance on biological data for large scale studies of ecological systems, we must begin to apply the quality assurance concepts that have been developed for environmental chemistry studies to biological sampling programs. We must be able to ensure the representativeness, comparability, and completeness of data sets. In addition, we must be able to ensure that we provide data of sufficient precision and accuracy to establish reliable estimates of status, as well as testing hypotheses regarding changing condition or trends in ecosystems. Only by applying QA/QC concepts to biological data collection efforts will we be able to produce results that will meet our program needs now and in the future.

Acknowledgements

We would like to gratefully acknowledge the support and assistance provided by Daniel T. Heggem during the preparation of this manuscript. This work was supported by the U.S. Environmental Protection Agency under Contract Number 68-03-3249 to Lockheed Engineering and Sciences Company.

References

Chapman, P. M., R. N. Dexter, and E. R. Long. 1987. Synoptic Measures of Sediment Contamination, Toxicity, and Infaunal Community Composition (the Sediment Quality Triad) in San Francisco Bay. Marine Ecol., 37:75-96.

La Point, T. W., S. M. Melancon, B. P. Baldigo, J. J. Janik, and M. K. Morris. 1983. Investigation of Methods for Site Specific Water Quality Assessment: Prickly Pear Creek, Montana. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada. 197 pp.

Mitchell-Hall, T. E., A. C. Neale, S. G. Paulsen, and J. E. Pollard. 1989. Eastern Lake Survey - Phase II: Quality Assurance Report. EPA/600/4-89/029. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada. 244 pp.

Peck, D. V., J. L. Engels, K. M. Howe, and J. E. Pollard. 1988. Aquatics Effects Research Program Episodic Response Project: Integrated Quality Assurance Plan. EPA/600/X-88/274. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish. EPA/444/4-89-001. U.S. Environmental Protection Agency, Office of Water, Washington, D. C.

Pollard, J. E. and D. T. Heggem. (Submitted.) Quality Assurance Sample Design for Large and Small Environmental Water Quality Surveys: A Look at the Past and Future. Submitted for publication in the symposia proceedings from the Conference on Environmetrics, Cairo, Egypt, April, 1989.

Pollard, J. E. and S. M. Melancon. 1984. Field Washing Efficiency in Removal of Macroinvertebrates from Aquatic Vegetation Mats. J. Freshwater Ecol., 2(4):383-391.

Stanley, T. W. and S. S. Verner. 1983. Interim Guidelines and Specifications for Preparing Quality Assurance Plans. EPA/600/4-83-004. U.S. Environmental Protection Agency, Washington, D. C. 29 pp.

Taylor, J. K. 1987. Quality assurance of chemical measurements. Lewis Pub., Inc., Chelsea, Michigan. 328 pp.

U. S. Environmental Protection Agency. 1978. Quality Assurance Guidelines for Biological Testing. EPA-600/4-78-043. U.S. Environmental Protection Agency, Environmental Monitoring and Systems Laboratory, Las Vegas, Nevada. 474 pp.

INTERCALIBRATION IN THE FRAMEWORK OF AN INTERNATIONAL MONITORING PROGRAMME

Merete Johannessen
Norwegian Institute for Water Research (NIVA)
P.O. Box 69, Korsvolol
0808 Oslo 8, Norway

Abstract

Experience from intercalibration in the International Cooperative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes with results from evaluation of monitoring data from Norway is discussed. The purpose of the programme is to use monitoring data to follow trends and changes in development of acidification or restoration, and to compare results from catchment studies across national boundaries.

The chemical programme is focussed on good laboratory practice. The programme also has to face the fact that strict rules regarding instrumentation and methods would make participation difficult for the member countries as the programme is based on national ongoing monitoring efforts and facilities available in each country.

With regard to biological monitoring, the problem of intercalibration has been minimized by the way data are treated and evaluated. In Norway, biological monitoring has become rather straight forward due to good knowledge of the geographical distribution of the fauna used for monitoring. Further, the water quality of the monitored areas is similar with regard to level of nutrition and content of dissolved salts. In these soft waters some climatic differences do exist, but the resulting effects can easily be accounted for when assessing presence or absence of species. The specific and easily detectable effects of acidification on the faunal composition should always be put in focus.

We are now addressing the question: What can these data be used for? The answer is that they will be used to identify a state of acidification. A simple model for estimation of an acidification score, based on the fauna, has been established.

We have accepted the fact that a systematic and detailed analysis of the fauna will not be achieved in most cases. However, generally accepted, easily identified, acid sensitive and tolerant species with well known distributions should be used. This will avoid uncertainty with regard to taxonomy and species geography. The philosophy is to deal with simple biotic indexes and fauna that are well known with respect to

acidity. This limited analysis normally will be sufficient to state a level of acidification.

Regarding the chemistry, a running intercalibration programme has been established based on the "Round Robin" concept. A quality assurance programme must address the system of sampling and calibration as well as chemical measurement. This wide definition is difficult to handle, but by visiting different participating labs we hope to enhance good laboratory practices. Four intercalibrations have been performed within this program so far. There have been good results for most major ions with a few exceptions. However, the exceptions are pH and alkalinity, which are of major importance in evaluation of the state of acidification. The next intercalibration will try to sort out what discrepancies are due to different methods, calibration practice and laboratory routine.

Whenever data are evaluated problems of intercalibration must be taken into account. However, when chemistry and biology point in the same direction, is that a coincidence? Field experiments from Norway show that acid sensitive invertebrates become established in watersheds during years of liming while sensitive species disappear during reacidification. This leaves us with the conclusion that the biological and the chemical results verify each other.

Introduction

In 1985 the International Co-operative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes was established under the auspices of the United Nations Economic Commission for Europe (UN/ECE). The Programme Centre is at the Norwegian Institute for Water Research (NIVA), Oslo. The University of Bergen (Dr. Gunnar Raddum) is responsible for the biology part of the programme in close cooperation with the programme centre at NIVA.

The reason for establishing this international programme was the need for improving our common understanding of the effects of acid rain on biota and freshwaters. The programme aims to harmonize, if possible, the different national programmes and to strengthen the basis for data evaluation.

The international monitoring programme became fully operative in 1987 after finalizing the Programme Manual. This was approved by the Programme Task Force in October 1986. (NIVA, 1987a). Data quality assurance and data quality control are addressed as key issues in the manual.

First of all the manual states the goal of the whole programme as:

"The International Co-operative Programme for Assessment and Monitoring of Acidification of Rivers and Lakes will be designed to establish, on a regional basis, the degree and geographical extent of acidification of surface waters. The data collected should provide more complete information on dose/response relationships under differing conditions and correlate the changes of acidifying deposition with the physical, chemical and biological status of lakes and streams".

More specifically the objectives of the programme are defined as:

"The primary objective of the programme for monitoring the changes in lakes and streams caused by acidification is to define long-term trends and variations in the chemistry and biota of surface waters attributable to the deposition of atmospheric pollutants, particularly acidic deposition. "Long term" here refers to variations over periods of years.

- Encompassed in this general objective are a number of specific sub-objectives which can serve as guidelines for programme design. These are:

  a) To determine trends of changes in chemical and biological parameters of aquatic systems having similar buffering capacities (sensitivity to acidification) and hydrology, but subject to a range of atmospheric acidic deposition levels;

  b) To compare the chemical and biological responses of aquatic systems having differing buffering capacities, but subject to similar atmospheric acidic deposition levels;

- Where the opportunity and resources exist, results from short term events can be reported on a voluntary basis, the purpose being to define and measure the magnitude of episodes".

To promote harmonization of monitoring strategies and to enhance data assurance, recommended guidelines are given in the manual on the following items (NIVA, 1987a):

- Monitoring programme objectives
- Sampling site selection
- Sampling methods
- Sampling frequency
- Sample handling and preservation
- Sample documentation and history
- Monitoring determinands
- Analytical methods
- Required accuracy

- Programme quality assurance and analytical quality control
- Data management and reporting procedures.

Regarding quality assurance, the programme has to face the fact that strict rules regarding instrumentation and methods would make participation difficult for many member countries as the programme is based on ongoing national monitoring efforts and facilities available in each country. Regarding chemistry, a recommended analytical programme of two levels has been suggested (Table 1). However, additional components can be added when available. Methods are recommended and required accuracy given.

An intercalibration programme has been set up and run to control the analytical routines and to identify problem areas. The running intercalibration programme has been established based on the "Round Robin" concept. As we shall point out later on, an international intercalibration programme can only be a supplementary tool to the running quality assurance programme at each laboratory.

With regard to biological monitoring, the problem of intercalibration has been minimized by the way data are treated and evaluated. In Norway, biological monitoring has become rather straight forward due to good knowledge of the geographical distribution of the fauna used for monitoring. Further, the water quality of the monitored areas is similar with regard to level of nutrition and content of dissolved salts. In these soft waters some climatic differences do exist, but the resulting effects can easily be accounted for when assessing presence or absence of species. The specific and easily detectable effects of acidification on the faunal composition should always be put in focus.

We have accepted the fact that a systematic and detailed analysis of the fauna will not be achieved in most cases. However, generally accepted, easily-identified, acid sensitive and tolerant species, with well known distributions, should be used. This will avoid uncertainty with regard to taxonomy and species geography. The philosophy is to deal with simple biotic indexes and fauna that are well known with respect to acidity. This limited analysis is normally sufficient to state a level of acidification. The model focuses especially on sensitive invertebrate species. Each of the localities within a watershed is given a score (scale 0-1) depending on presence/absence of organisms indicating different levels of acidity.

The list of sensitive species used in the different categories in western Norway is given in Table 2. The tolerance limits are based on field observations, experiments, and data from the literature. They are valid only for softwater (specific conductance $< 30\,\mu S/cm$). The sample can consist both Surber and kick samples. It is the relative species composition that is of importance and not the quantity in this system. Raddum et al. (1987) and the manual (NIVA, 1987a) give additional information about sampling strategy and processing of data.

Table 1: Analytical programme suggested for the international monitoring programme (ammonium is a primary determinand in areas were this is important).

Primary (essential) determinands: *	
Anions:	alkalinity (or acidity), sulphate, nitrate, chloride, dissolved organic carbon
Cations:	pH(H^+), calcium, magnesium, sodium , potassium, total aluminum, ammonium
Physical properties:	temperature, discharge (rivers), specific conductance, water level.
Secondary (desirable) determinands:	
Aluminum speciation:	fractions of "labile" and "reactive" aluminum
Nutrients:	ammonium, total nitrogen, total phosphorus, dissolved oxygen, silica
Metals:	iron, manganese, cadmium, zinc, copper, nickel, lead
Other:	fluoride, colour

* If a complete primary analysis cannot be obtained, a determination of pH, temperature, specific conductance, alkalinity and streamflow or water level may be accepted by the Programme Centre as a minimal monitoring report.

It is important that this way of treating biological data has been decided upon in the manual. However, many countries from different areas may conduct more detailed biological programmes. These data are also transferred to the programme data base and can be used for additional comments. However, this most simple acidification score model provides a common means of treating data and evaluating data collected in different ways.

Table 2: **Acidification-sensitive invertebrate taxa used in monitoring of acidified watersheds in Norway.**

Category	Species/Group	Score
a	Gastropoda Crustacea: *Gammarus lacustris* *Lepidurus arcticus* Ephemeroptera: *Baetis* spp.	1
b	Cladocera: *Daphnia* spp. Ephemeroptera: *Siphlonurus* spp. *Ameletus inopinaus* Plecoptera: *Isoperla* spp. *Diura* spp. *Capnia* spp. *Leuctra fusca* *Arcynopteryx compacta** *Dinocras cephalotes** Trichoptera: *Apatania* spp. *Hydropsyche* spp. *Philopotamus montanus* *Lepidostoma hirtum* *Ityrichia lamellaris* *Glossosoma* sp.	0.5
c	Small mussels (Sphaeriidae)	0.25
d	No registration of above mentioned species or groups	0.0

* Rare species in Western Norway

The Programme Status

The programme became fully operative in 1987 after finalizing the programme manual in January 1987. The Programme Centre prepares annual data reports and conducts intercalibration exercises. Links are also being established with other ongoing and planned environmental networks in Europe and North America. The 1987 data report, the 1988 data report and three intercalibration reports are now available. At the moment, the international database at the Programme Centre at NIVA includes data and site information from 143 catchments in 12 countries; Belgium, Canada, Denmark, Ireland, Finland, Federal Republic of Germany, The Netherlands, Norway, Sweden, United Kingdom, United States of America and the Union of Soviet Socialist Republics. In addition, other countries participate in the annual Task Force meetings and are encouraged to join the programme. Table 3 shows the participation at the Task Force meeting in October 1989.

Table 3: Programme status as of the Task Force Meeting, October 1989.

Country	Delegates To Task Force Meeting	Official "member"	Partici-pating inter-calibr.	Data reported for 1988
Belgium	X	X	X	X
Bulgaria	X			
Canada	X	X	X	X
Denmark	X	X	X	
Ireland	X	X	X	X
Finland	X	X	X	X
France	X			
Germany FR	X	X	X	X
Hungary	X			
Italy	X		X	
The Netherlands	X			
Norway	X	X	X	X
Sweden	X	X	X	X
United Kingdom	X	X	X	X
USA	X	X	X	X
USSR	X			
SUM	16	10	10	9

The three-year report for the period 1987-89 will present the results of the programme to this point. This report will include the first evaluation of results. A draft outline of the programme report has been agreed upon in advance by the programme Task Force.

Running The Programme

The manual gives the platform of the programme. Attainment of the overall goal, however, depends on several quality assurance aspects that are specified only as recommendations in the manual. The Programme Centre performs several key tasks such as:

- Assure that a "broad" distribution of measuring sites are included, i.e., both with respect to geographical distribution and water quality.

- Work to achieve improvement regarding programme design and analytical work in each of the participating countries.

- Review discrepancies from the recommended methodology and to take this into account in the evaluation of data.

Relevant sampling site distribution

To assure a relevant distribution of sites mapping the geographical extent of acidification and sensitive areas of Europe and North America constitutes an important part of the monitoring programme. Maps of sensitive waters at the national level based on geology and soils maps, using various criteria for definition of sensitivity, have been combined into a map of water sensitivity for Europe (Skjelkvale and Wright, 1990).

High	:	High sensitivity
Grey	:	Moderate sensitivity
Dotted	:	Low sensitivity
White	:	Areas with inadequate or no information about the sensitivity

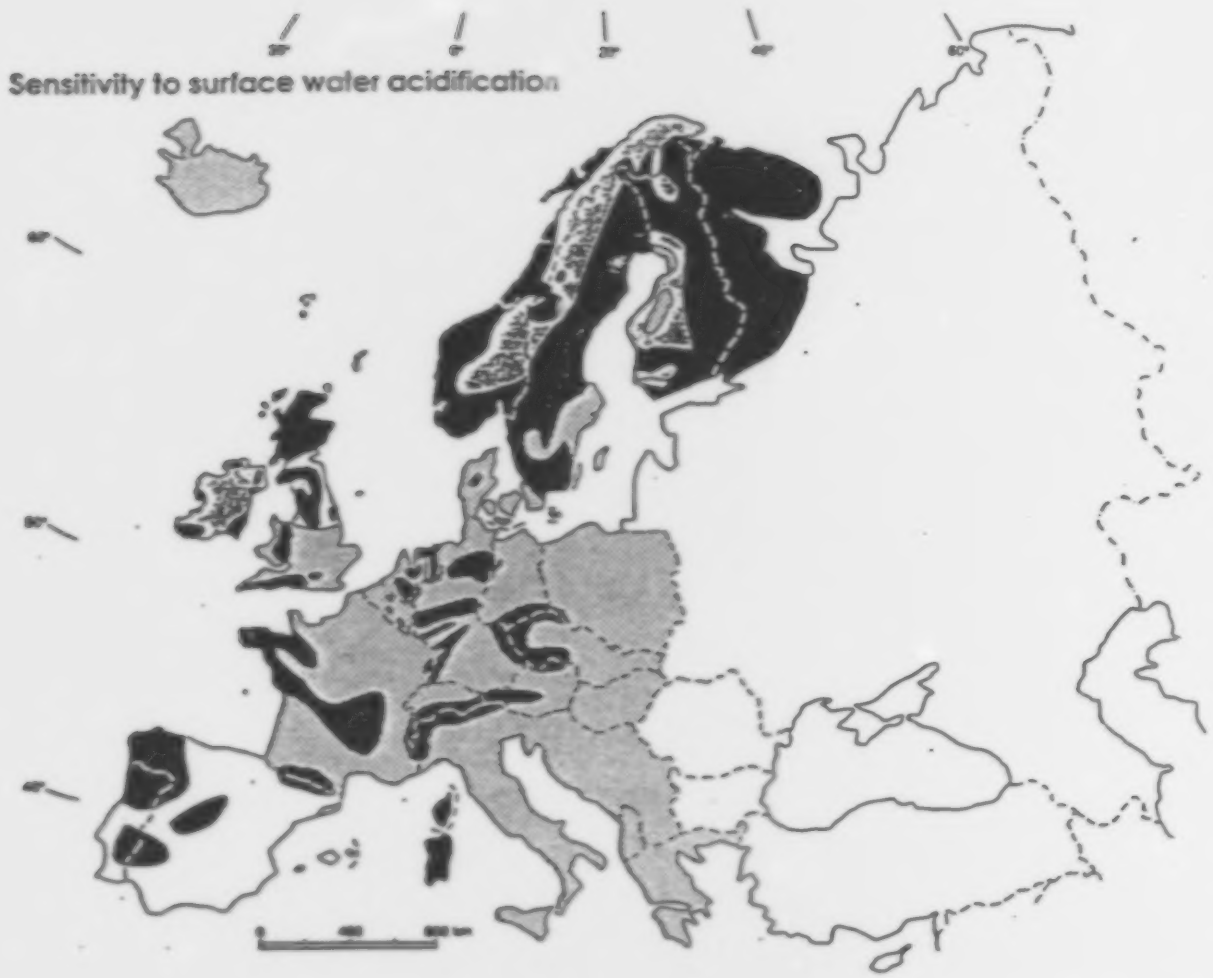


Figure 1a. Sensitivity map for Europe (from Skjelkvale and Wright, 1990).

International Co-operative Programme for Assessment and Monitoring of Acidification of Rivers and Lakes.
Sampling locations as of October 1989

Figure 1b. Monitoring sites in Europe as of October 1989.

This international programme relies entirely on national data collection/gathering programmes. At the moment, the sites are reasonably well distributed geographically (Figure 1b) and chemically (Figure 2). However, background stations in sensitive areas not receiving acid rain are few.

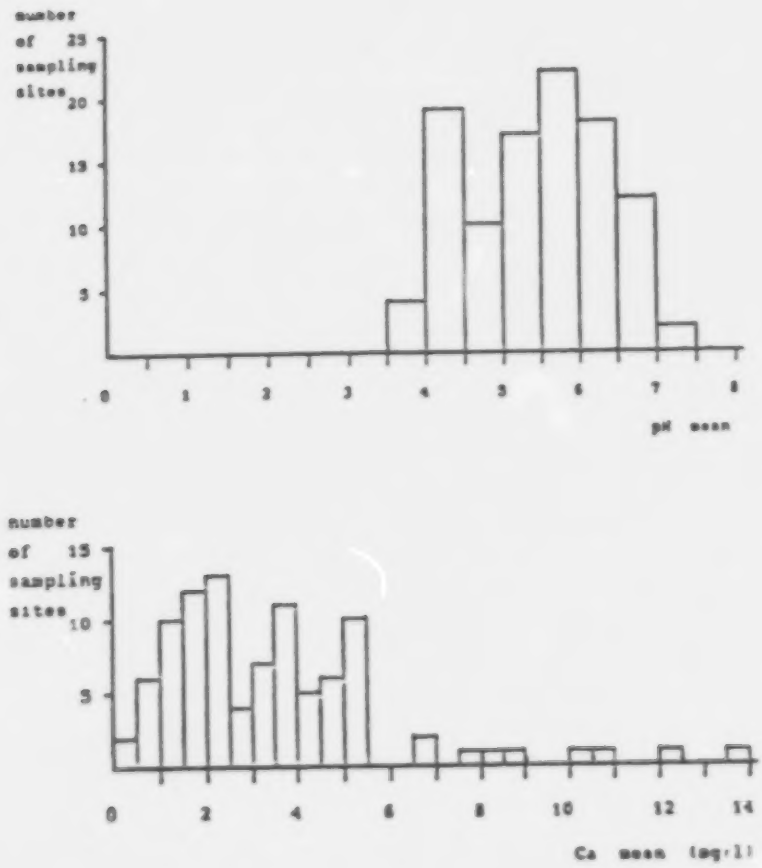


Figure 2. Frequency distribution of annual mean pH and annual mean Ca concentrations of the sites in 1987 (NIVA, 1989).

Continuous search for improvements

The Programme Manual states:

"In a co-operative programme, quality assurance must be applied to all phases from sampling procedures and analytical determinations to data editing and archiving. This quality assurance is extremely important in assessing comparability of the data and in eventual attainment of the

programme objectives. It must be achieved through development of standardized guidelines and procedures, strict co-operative adherence and thorough documentation.

An extremely important part of the overall quality assurance is the analytical quality control, whose function is to assure accuracy and comparability of chemical analyses carried out by the various participating laboratories. Both in-laboratory and between-laboratory quality control procedures are vital to the success of co-operative monitoring programmes".

The procedures employed are based on the "Round Robin" concept and Youden's methods (Youden, 1959; Youden and Steiner, 1975). Samples prepared by the Programme Centre are targeted to test for bias in analyses of the principal determinants of the programme. Participating laboratories are assessed on their reported determinations in relation to the other participants.

Twice a year two synthetic water samples with different concentrations of the particular ions in question are distributed. Periodic assessment reports provide evidence of analytical bias. These reports are of primary use to initiate corrective actions that will eliminate the bias. In an international programme based on national contributions of data, an inter-laboratory control programme can be undertaken to identify systematic or random errors. Here is an example to illustrate the situation.

From the first intercalibration run with synthetic samples we found that for sulphate and calcium an improvement of the analytical methods at some of the laboratories is required to yield better comparability between the laboratories in this group. Especially at lower concentrations, some laboratories obtain results that deviate too much from the true value. However, pH and conductivity results were acceptable (NIVA, 1987 b).

The second intercalibration (Hovind, 1988) showed some improvements. For conductivity, nitrate, sulfate and sodium, all the results were within the limits based on the general target accuracy. For chloride, calcium, magnesium and potassium, one or two laboratories reported values outside the general limit. By some improvement of the routine analytical methods, these laboratories should be able to attain acceptable results. In certain cases, data quality interpretation is based on narrower limits (the so-called special limits). Here for pH and alkalinity the picture is more serious. It is possible that normalization of analytical methods could improve the result. Indeed the Programme Centre has asked most of the laboratories to make such improvements.

The third intercalibration (Hovind, 1989) dealt with organic carbon and aluminum fractions. Based on the general target accuracy of 20%, all results were acceptable for dissolved organic carbon. With regard to aluminum fractions, the great discrepancy in methods used is reflected in the range of results. To compare these parameters the same method must be used.

Compilation of the different alkalinity values are of great importance in the understanding of the results. Intercalibrations so far have not given satisfactory results. The fourth intercalibration (samples circulated in December 1989) addresses this problem in particular and was designed to sort out systematic from random deviations.

No international intercalibration programme can substitute for daily internal laboratory control procedures. This international programme relies upon the participants' own interest in improving quality control. Information, meetings, publicity, and laboratory visits are all possibilities to be used in this respect.

Deviation from Recommendations

The Programme Centre gives an overview of deviation from the recommended situation in several aspects . First intercalibration results always accompany the data in the annual reports. The annual report of 1988 also gives an overview of the collected site selection criteria and conditions. These data are entered in the data base so as to show how site selection criteria are fulfilled.

As shown, the idea is to enhance good scientific practice in the field and in the laboratory so as to maximize quality assurance. However, the possible difference between recommended practice and actual practice in each case has to be accounted for in evaluation of the available data. For example, the latest intercalibration attempts to sort out deviations due to different analytical procedures, with the intention to derive standardized corrections, if possible.

The programme strategy is to utilize the data base so as to evaluate theory and models based on the most extensive studies, to do pilot studies and to apply the models on research sites with more limited data available. Also, data from other research programmes can be used to establish methods. In particular , older data from research sites may be entered into the database and used.

It has also been agreed that data from the monitored sites prior to the start of the international program may enter the data base.

The combination of chemical measurements with biological investigations addressing species sensitive to acid conditions is the strength of this monitoring programme.

Examples of Pilot Results

An investigation of 71 localities with regard to invertebrates and water chemistry in Norway offers a good example of data treatment. The localities are in the softwater area of Western Norway and at the moment the conclusions are only valid for these areas. In acidification scoring, the unchanged fauna give score 1, while the more affected areas have scores down to 0. In this particular example, faunal assemblages with scores 1, 0.5 and 0 are used and connected to water quality.

Figure 3 shows the relative distribution of "lake-acidification score" for a number of lakes belonging to each pH interval measured at the time of biological sampling. This illustrates the connection between a single pH measurement and the fauna at the spot. However, using the calculated ANC value (ANC is defined as acid neutralizing capacity) instead of pH, a narrower band is obtained (Raddum and Lien pers.comm.).

The use and value of biological indicators depends on possible recolonization during periods of improved water quality. The Norway liming experience might confirm this. One example from the River Audna is shown in Figure 4. The acidification index from River Audna is improved from almost 0 to 0.7 during liming activities, starting in spring 1985. During the 5 years of liming, pH has been kept at 5.5 to 6.5, almost one pH unit higher than prior to treatment. The most sensitive species Beatis rhodani did not occur until autumn 1987. The fish catch improved, especially for migratory brown trout (Salmo trutta). Salmon have also re-established in the area after decades of absence. This verifies the sensitivity and use of benthic fauna as a good indicator of water quality change.

The use of acidification scoring techniques must take into account different faunal compositions in areas with quite different water quality and climate. Knowledge of the fauna from areas with extended biological distributions can be valuable for nearby areas in the same climatic zone across national boundaries.

Whenever data are evaluated problems of intercalibration must be taken into account. However, when chemistry and biology point in the same direction, is that a coincidence? Field experiments from Norway show that acid sensitive invertebrates become established in watersheds during years of liming, while sensitive species disappear during reacidification. This leaves us with the conclusion that the biological and the chemical results verify each other.

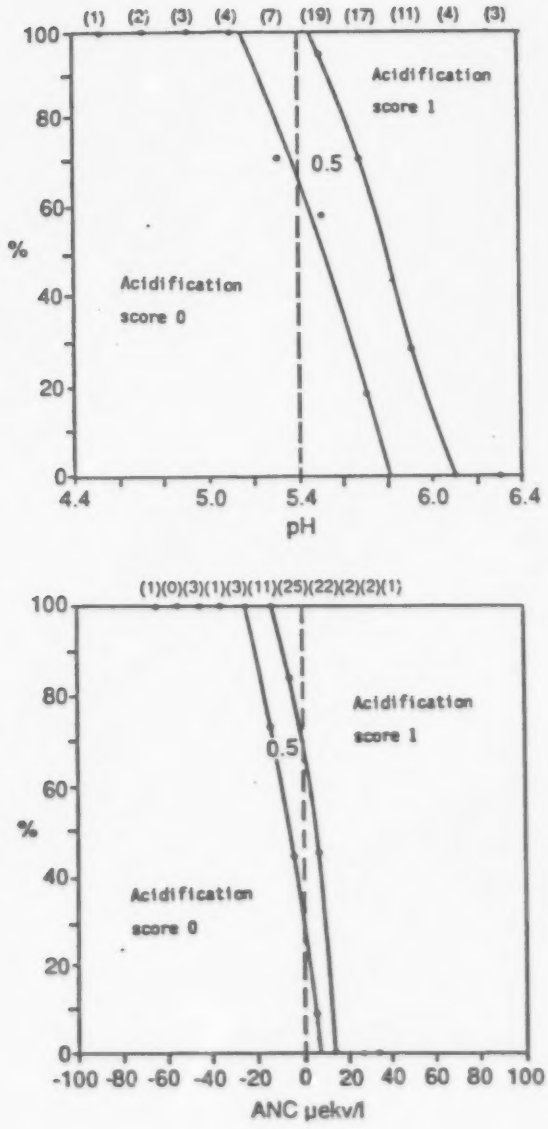


Figure 3. Acidification score, percent of localities with a given score at each pH interval.

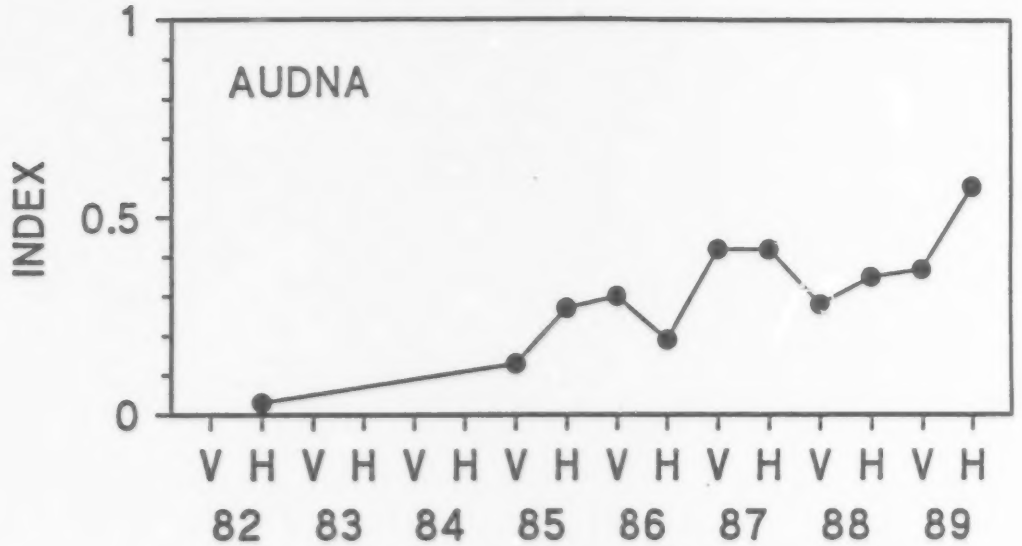


Figure 4. Acidification index of River Audna. Improved water quality due to liming, starting spring 1985.

Final Remarks

Establishing and running an international monitoring programme calls for special attention to several points. One major advice is to start with a specific but simple strategy regarding chemical and biological measurement. Secondly, we strive for improvements all the time. Within the International Co-operative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes, direct interest from the scientists involved is the very best way to collect data for a reliable and extensive data base, and to be able to utilize the data in an up-to-date scientific manner.

References

Hovind, H. 1988. Intercalibration 8802: pH, K_{25}, HCO_3, NO_3, SO_4, Cl, Ca, Mg, Na, K. International Co-operative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes. NIVA, 1988. 40 pp.

Hovind, H. 1989. Intercalibration 8903: Dissolved organic carbon and aluminium fractions. International Co-operative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes. NIVA, April 1989. 27 pp.

NIVA. 1987a. Manual for Chemical and Biological Monitoring. International Co-operative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes. NIVA March, 23 p.

NIVA. 1987b. Intercalibration 8701: pH, Ks, SO_4, Ca. International Cooperative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes. 28 pp.

NIVA. 1989. Data report 1987 and available data from previous years. International Co-operative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes. NIVA April, 475 pp.

Raddum, G., Fjellheim. A. and Hesthagen T. 1988. Monitoring of acidification by use of aquatic organisms. Verh. Internat. Verein. Limnol. 23, 2291-2297.

Skjelkvale, B.L. and Wright, R.F. 1990. Overview of areas sensitive to acidification: Europe. NIVA (in press).

Youden, W.J. 1959. Graphical Diagnosis of Interlaboratory Test Results. Industrial Quality Control. pp 15-24.

Youden, W.J., Steiner, E.H. 1975. Statistical Manual of the Association of Official Analytical Chemists. Statistical Techniques for Collaborative Tests. Arlington.

THE ROLE OF QUALITY ASSURANCE IN CONDUCTING TERRESTRIAL ECOSYSTEM MONITORING

R.P. Breckenridge*
Center for Environmental Monitoring and Assessment
Idaho National Engineering Laboratory
Idaho Falls, ID.

Abstract

Quality assurance in ecosystem monitoring helps reduce the variability associated with human factors, and increase the probability that if the work is duplicated by other scientists, the results will be comparable. This paper addresses quality assurance issues related to program design, field work, and subjective measurements in terrestrial monitoring. The purpose of this paper is to provide topics for discussion for participants at the Ecological Quality assurance Workshop, and is not intended to be an exhaustive study of these topics.

Introduction

Ecosystems are dynamic, with a high number of complex processes operating at any given moment. Uncertainties in ecosystem monitoring stem from both natural variability and human factors. These could include anything from daily fluctuations in metabolic processes, to the use of contaminated sampling equipment in the field. All this makes the development of a useful quality assurance program challenging.

Quality assurance helps reduce the variability associated with human factors, and increase the probability that if the work is duplicated by other scientists, the results will be comparable. The characteristic of comparability reflects both internal consistency of measurements and expression of results in units consistent with other organizations reporting similar data (DOE, 1989).

---

* in collaboration with G.B. Wiersma and G.L. Olson, Center for Environmental Monitoring and Assessment, Idaho National Engineering Laboratory, Idaho Falls, ID; and K.B. Jones, Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency, Las Vegas, NV.

In order to establish a sound quality assurance program, all phases of the monitoring activity must be addressed: program design, field work, laboratory procedures, data interpretation, and reporting. This paper briefly discusses quality assurance issues related to program design, field work and subjective measurements in a terrestrial monitoring program. This paper is not intended to be an exhaustive study on these topics, but rather to introduce topics for discussion amongst the workshop participants. This paper reflects the opinions of the authors and these are preliminary.

Program Design

Quality assurance for ecosystem monitoring starts with the program design. The National Research Council (1990) published a conceptual framework to help guide the development of sound monitoring programs. This framework can be used to provide "a logical and scientifically based means of linking technical decisions about monitoring design to the information needs of the decision-making process" (NRC, 1990). Figure 1 shows the elements of designing and implementing a monitoring program.

This figure illustrates the need for posing pertinent questions at the right junctures; then refining, restating or rethinking the objectives of the project to obtain pertinent answers. The iterative process helps assure that resources spent on data will answer questions that are asked. This assurance, in the form of answerable questions, must be established before the monitoring design can take form. The U.S. Fish and Wildlife Service is currently in the process of revising its National Contaminant Biomonitoirng Program because of a "no" answer obtained in response to the question "Can changes be detected?" (see Figure 1). This is because the pesticides the program was originally designed for have been banned and chemical residues have, in many cases, dropped to or below detection limits. As a result of this scrutiny, the Service has discontinued monitoring of duck wings and starlings for DDT, and is now looking at other possible biological approaches that could produce a "yes" to this question. Such scrutiny is not only required at the outset of a program, but is required routinely to determine that the objectives are being met and are current and the program is producing useful results.

Sampling design, or re-design, can only be developed after the expectations and goals of the program are clearly defined. The expectations and goals provide the framework for establishing data quality objectives, and making decisions about the sampling design. The sampling design will largely be driven by the level of confidence the researchers require in order to meet the data quality objectives. For regulatory monitoring programs, it may be necessary to establish more stringent requirements than for early-warning indicator monitoring, where false positives are acceptable (Kelly and Harwell, 1988). The number of samples to take will depend on the data quality objectives, and on the type of sampling to be done (e.g., composite, two-stage, grab), sample design (simple or stratified random, systematic), whether kriging is involved, etc.

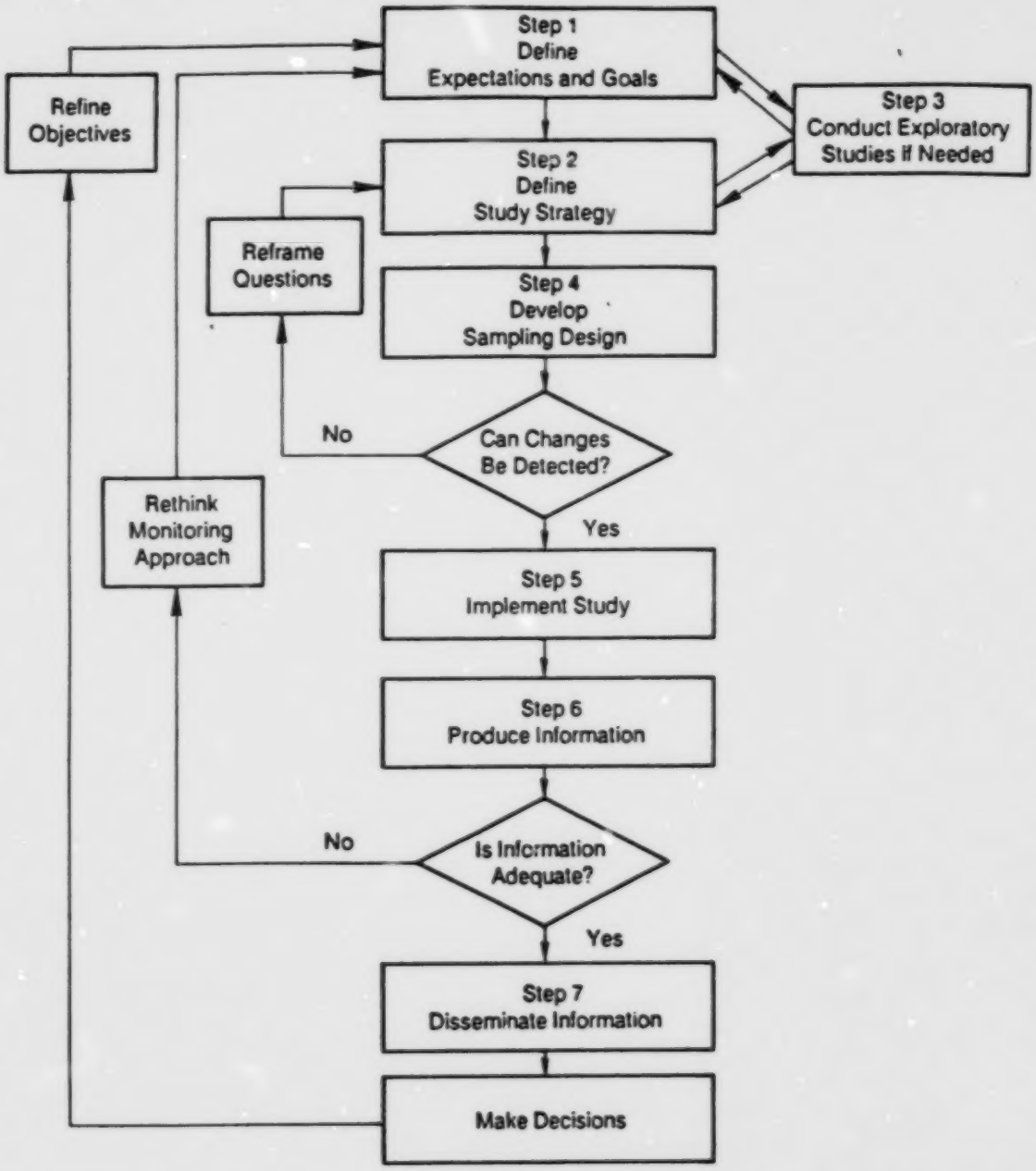


Figure 1. The elements of designing and implementing a monitoring program (Source: NRC 1990).

Also guiding the development of the sampling plan will be: the cost-effectiveness of alternative sampling designs, patterns of environmental contamination and variability, and practical considerations such as convenience, site accessibility and availability, security of sampling equipment, and political considerations (Gilbert, 1987).

Field Sampling

Following are some factors to take into consideration when designing a field quality assurance program.

Sample Timing

The object of considering sample timing is to minimize the variability of the data yet permit objectives to be met. The best time to collect a sample is often referred to as the index period. This is an important issue when collecting remote or ground verification data. The index period can change with the time of year or day and is a critical value to assess when considering data variability. By considering the following issues, some QA concerns can be allayed.

1. Seasonal influence and multiple measurements over the year--There are numerous instances when time of the year is critical to the objectives of the project. Vegetation assessments, for example, are probably best made in the spring or summer. Sample timing for soil samples, however, is generally less important since fluctuations in the soil chemical status are more stable across seasons. Whether such considerations are important may be determined through literature searches on similar ecosystem monitoring studies.

2. Time of day--Photosynthetic processes and animal activity are influenced by the time of day and thus need to be evaluated.

3. Meteorologic activity--Several meteorologic parameters influence the amount of variability in monitoring results. For example, a study done by Miller (1963) showed that in general, the variation in the amount of nutrients that reached the ground from leaf drip generally were consistent with the volume of leaf drip, except in cases of high rainfall. Although field situations do not necessarily allow for the minimization of such variability, significant sources of variance should be recorded.

4. Ecosystem cycles or activity--A familiarity with the life cycles of the target population is critical in monitoring some ecosystem parameters. Leaf-out stage and flowering may be significant considerations when monitoring vegetation. Studies of brood success for each species need to be carefully coordinated with the appropriate time of year.

Site Selection

Ideally, site selection criteria would be concise enough so that if two researchers were to follow them, each would choose similar locations. The criteria should minimize the amount of subjectivity that enters into the site selection process.

In choosing a site to monitor, the study and data quality objectives need to be examined first. For example, prior to choosing the Wind Rivers as a site for acid deposition research, Bruns et al. (1987) identified several criteria applicable to their objectives. These were: ecosystem components, pollution sources, site logistics, and available support (Table 1). These criteria, coupled with a usable heuristic tool (Kepner and Tregoe, 1981) led to a logical, well documented selection process, rather than a haphazard one.

In order to ensure comparability between sites, selection criteria should be carefully chosen. For example if the terrestrial area to be monitored is a remote site, factors such as: influence from global and local pollution sources should be assessed and kept to a minimum; the area should be located in an area with minimum economic development for the coming decades; the core study area should be restrictive against human intervention; a relevant research institute should be located nearby; and, have baseline data available for the site with particular emphasis on long-term (30-50 years) metecrological records (Wiersma, 1983).

Site selection criteria will depend on a number of factors, including the scale of the program or project. For monitoring a landscape or watershed, vegetation type, slope, aspect, drainage class and soil type should be taken into consideration when selecting a site. For monitoring for regional or national representation, geology, physiography, vegetation, precipitation and moisture regimes, and climate should be considered.

Sample Collection

Natural sample variability can be reduced through various sampling methods. The most common way to reduce sample variability is to composite. Soil or vegetation samples from different corners of a grid are frequently composited to obtain an average for the grid square. Another way to reduce variability is to take a sufficiently large homogenized sample from which to draw a smaller subsample. These methods work well for small scale or site-specific assessments; when, for example the whole site can be gridded.

TABLE 1. CRITERIA AND SITE EVALUATION FOR ACID DEPOSITION RESEARCH

	Criteria	(Wt 1-5)[a]	Winds[c]	Uintas[d]
Ecosystem Components				
1.	High elevation granitic basin	5	25[b]	20[b]
2.	Lakes with fishery potential	4	20	16
Pollution Sources				
3.	Upwind acid sources	5	25	25
4.	Potential to receive acid deposition	5	20	25
Site Logistics				
5.	INEL access	3	9	6
6.	Winter access	4	16	12
7.	Summer access	3	16	12
8.	Local public influence	4	16	8
Available Support				
9.	Near NADP site	2	10	2
10.	Other data and research	3	9	3
11.	Power sources	5	20	10
12.	Field manipulations	5	15	20
	TOTALS		201	159

a. Weighting factors
b. These values are based on relative scores at each site multiplied by the appropriate weighting factor in column one.
c. Proposed site on the west side of the Wind River Mountain Range in Wyoming.
d. Proposed site in Uintas Mountains of Utah.

However, even with composite samples, the variance can vary greatly from year to year (Table 2). Data collected from a remote forest site in Southern Chile (Wiersma et al. in press) show, in general, a high coefficient of variances (CV) for the natural elements in vegetation as compared to NBS standards (Table 2). It also shows that depending on the element there can be a significant difference between different natural vegetative types. For example, CV for phosphorus in moss were as high as 152.2% compared to 41.5% for lichens during the same collection periods. As long as this is noted and accounted for in design, the monitoring program can be successful. Also, this type of information provides guidance for modification of the monitoring program as needed.

Subjective Measurement and Training

Subjective measurements include those that rely on subjective estimates. These systems are relied on heavily in ecosystem studies (Barbour et al., 1987), such as vegetation cover percentages and species composition. The problem with them is that there is a high degree of variability between samplers. An estimate of vegetative cover made by a plant ecologist will often differ from the estimate on the same plot of land by a technician.

One way to reduce the subjectivity from field estimates is by using standardized field sheets that provide multiple-choice type answers in addition to spaces for other field notes. For example, the soil description forms used by the USDA allow soil scientists with little training to produce fairly accurate soil descriptions. These forms include spaces for all relevant information, such as structure, texture, soil color, horizon depth, etc. There are standard ways to report each of the properties, e.g., the size, strength and shape of the soil ped are all described in the soil structure section. A soil scientist making a soil profile description without a soil description sheet is much more likely to omit pertinent information than the soil scientist that has the sheet. Similar sheets should be required for all ecosystem studies that require documentation of data from the sampler. These should provide rules of thumb commonly used to make estimates. The sheets should be in a form that are easily entered into a computer database. QA checks must be designed for this part of the monitoring process also.

Training is vital to the success of the quality assurance program that addresses subjective measurement systems. The U.S. Forest Service, for example, is addressing this concern now and working on standardizing data collection so that data can be used in Geographic Information Systems. However, for this to be successful, all employees involved in the monitoring need to receive the same level of training and be checked by qualified field auditors. Management must accept the cost and time required to complete the training.

TABLE 2. CV VALUES FOR ELEMENTS ANALYZED BY SPARK SOURCE EMISSION SPECTROSCOPY (ALEXANDER AND MCANULTY, 1981)

Vegetation Type	P (%)	K (%)	Ca (%)	Zn (%)
Moss	48.1-152.2	13.7-55.9	16.0-78.2	28.0-98.2
NBS Standard Tomato leaf	2.9-18.9	7.1-14.1	2.1-18.6	15.6-20.3
Lichens	23.8-41.5	12.2-29.6	42.4-82.4	17.8-32.9
NBS Standards Tomato leaf	2.5-5.1	7.0-25.6	2.2-5.6	9.1-17.0
Lettuce leaf Standard	8.7-19.4	2.6-7.4	3.3-5.9	4.5-19.8

Data for samples collected during: March 84, November 84, March 85, April 88.

Conclusions

There are numerous challenges to obtaining quality-assured data in environmental monitoring due to the high degree of natural variability in ecosystems and human error. Quality assurance programs in terrestrial ecosystem monitoring helps reduce variability associated with human factors, and increases the probability that the results will be comparable with other studies. Quality assurance in monitoring starts with sound program design, including the formulation of specific questions. With answerable questions in place, the sampling design can be developed. The sampling design must take into consideration the data quality objectives, which will guide the formulation of sample numbers and quality assurance needs. Sample timing, site selection, and sample collection concerns also need to be taken into consideration. Training is necessary to minimize error and ensure that the samples taken will be taken in the correct manner.

References

Barbour, Michael G., Jack H. Burk, and Wanna D. Pitts. 1987. Terrestrial Plant Ecology, Benjamin/Cummings Publishing Co. Inc., Menlo Park, CA.

Bruns, D. A., G. B. Wiersma, G. J. White, C. S. Staley, R. D. Sanders, R. D. Rogers, and D. Miley. 1987. Preliminary Results and Research Protocol for Environmental Functioning and Response to Energy Related Emissions, EG&G Idaho, Inc., Internal Technical Report.

Bruns, D. A. and G. B. Wiersma. 1988. Research plan for integrated ecosystem and pollutant monitoring at remote wilderness study sites, INEL Informal Report EGG-EES-7951.

Gilbert, Richard O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Company, New York.

Kelly, John R., and Mark A. Harwell. 1988. Indicators of Ecosystem Response and Recovery, In Ecotoxicology: Problems and Approaches. Springer-Verlag.

Kepner, Charle H., and Benjamin B. Tregoe. 1981. The New Rational Manager. Princeton Research Press.

Miller, R. B., Soil Bureau, Department of Scientific and Industrial Research, Taita. 1963. Plant nutrients in hard beech: The cycle of nutrients. New Zealand Journal of Science.

National Research Council. 1990. Managing Troubled Waters: The Role of Marine Environmental Monitoring. National Academy Press, Washington, D.C.

Oakes, T. W. 1982. Quality Assurance in Environmental Measurements. Department of Environmental Management, Oak Ridge National Laboratory.

U. S. Department of Energy and Idaho National Engineering Laboratory. 1989. Draft. Sampling and Analysis Data Document for: Stanford Linear Accelerator Center, Lawrence Berkeley Laboratory. DOE Environmental Survey.

Wiersma, G. Bruce, D. A. Bruns, C. Boelcke, C. Whitworth, and L. McAnulty. 1990. Elemental Composition of Mosses from a Remote Nothofagus Forest Site in Southern Chile. Chemosphere (in press).

QUALITY ASSURANCE ASPECTS OF THE JOINT USA - CANADA SUGAR MAPLE DECLINE PROJECT

W.G. Burkman*
NSI Technology Services Corporation
Radnor, PA 19087

Abstract

The North American Sugar Maple Decline Project was implemented in 1988 in response to concerns about the condition of sugar maple (Acer saccharum Marsh.) in the United States and Canada. Data collection for the project involved at least 15 federal, state, and provincial crews from seven states and four provinces. The response measurements included four subjective measurements of tree crown conditions: branch dieback, foliage transparency, foliage discoloration, and dwarfed foliage.

Developing the quality assurance program for this project presented some major challenges. To ensure the best possible data quality the following quality assurance activities were implemented: 1) detailed documentation of the measurement systems; 2) a comprehensive training session for all participating crews and other project personnel; 3) use of independent audits to review field crew procedures; 4) remeasurement of a subset of plots to assess the precision of the data; and 5) use of the remeasurement data, audit information, and training session evaluations to improve the project over time.

The effectiveness of the quality assurance program in this project is substantiated by the remeasurement precision of overall tree crown conditions for the 1988 field season: branch dieback - 95.6%; foliage transparency - 90.9%; foliage discoloration - 91.3%; and dwarfed foliage - 88.3%. The percent of the trees remeasured was 5.39% and 8.32% in 1988 and 1989, respectively. The identification of methodology interpretation problems in foliage discoloration and dwarfed foliage led to an improvement in agreement percentages in 1989. All tree crown condition ratings agreement percentages increased in 1989 (as compared to 1988): branch dieback, +1.1; foliage transparency, +0.9; foliage discoloration, +7.8; and dwarfed foliage, +10.5%.

---

* in collaboration with I. Millers, USDA Forest Service, Durham, NH 03842; and D. Lachance, Forestry Canada, Laurentian Forestry Centre, Ste.-Foy, Quebec G1V 4C7.

The results from this project show that quality assurance activities can quantify the quality of data for subjective measurements. In addition, quality assurance procedures can improve data quality over time.

Introduction

The implementation of quality assurance (QA) has been shown to be important in field research projects and ecosystem research programs (Medlarz, 1988; Cline and Burkman, 1989). Although QA has been implemented on an individual project basis, the use of QA in regional research programs has been limited. This area of QA implementation will become very important as new, major monitoring programs such as the Environmental Monitoring Assessment Program are implemented in the U.S. This report describes the implementation of QA in a large, regional project. A basic understanding of the North American Sugar Maple Decline Project is necessary so that interpretation of the QA components can be related to the project objectives. Therefore, a short overview of the project will be presented.

Overview of the North American Sugar Maple Decline Project

Concern about sugar maple (Acer saccharum Marsh.) decline in the U.S. (Millers et al., 1989) and Canada (Lachance, 1985) resulted in a cooperative agreement for a joint research project between the USDA Forest Service and Forestry Canada to investigate whether a decline existed and, if so, the extent and causes of the decline. The project was funded through the Eastern Hardwoods Research Cooperative of the Forest Response Program, which operated as part of the National Acid Precipitation Assessment Project.

The objectives of the North American Sugar Maple Decline Project were to:

1. Determine the rate of change in sugar maple tree crown condition ratings from 1988 through 1990.

2. Determine if the rate of change in sugar maple tree crown condition ratings are different between:
   a. sugarbush and undisturbed forested stands,
   b. various levels of initial stand decline condition,
   c. various general levels of pollution measured as wet deposition.

3. Determine the possible causes of sugar maple decline and the geographical relationship between causes and extent of decline, if the decline existed.

A jointly developed work plan and a field manual documented instructions for plot selection, establishment, and data collection. This manual was used by all field crews. Additional supplements included a QA plan and a data analysis and interpretation plan.

Minor changes in methodology were implemented after the 1988 field season based on input from project coordinators and the remeasurement data.

A total of 166 clusters of 5 plots each were located from Wisconsin to Nova Scotia and represented a wide range of pollution levels, initial stand decline conditions, geographical areas, and physical and biological conditions in southeastern Canada and the northeastern / northcentral U.S. One-half of the clusters were located in sugar maple stands managed as sugarbushes (used for the production of maple syrup) and one-half in undisturbed forest stands.

The number of established clusters and the cooperating agencies responsible for the surveys in the clusters are listed in Table 1. Because of the large number of cooperating agencies, at least 15 field crews participated in actual data collection.

The project data was collected in three major categories: site descriptions, stand descriptions, and tree measurements. A complete listing of project methods is contained in the Field Manual for the North American Sugar Maple Decline Project (Millers et al., 1990).

Site descriptions included a general description which covered an area approximately 10 km^2 around the cluster, and a local description at each plot. The site data included location, regional physiography, and local physiography. Regional physiography included information regarding terrain, major watershed, weather data, and soil series. Usually the data was obtained from published materials.

Project data was collected that described general stand characteristics. A stand was defined as a similar portion of a hardwood forest in which more than 50 percent of upper story (dominant and codominant) trees were sugar maples 50 to 150 years old. Stand information included descriptions of past disturbance and vegetation. Disturbance information described type (logging, grazing, and sugaring) and severity. Vegetation information included data on crown closure and structure, site quality, and stand age. The two major types of stands, based on management history, were sugarbush and undisturbed forest and defined as:

1. Sugarbush - A hardwood stand with some evidence of sugaring in the last 5 years and at least a quarter of the trees with tapholes.

2. Undisturbed Forest - A hardwood stand with no evidence of disturbance in the last 5 years, such as sugaring or logging. The stand should be at least 2 ha in area, but preferably larger.

The data category for tree measurements was subdivided into two groups, descriptive and response data. Descriptive information included species, diameter at breast height (dbh), crown position, a general vigor rating, and both location and description of bole

defects for all trees on a plot. This information, with the site and stand data, was collected to provide background information for the analysis of response data.

Response data included information from the crown rating system. In 1988, response data was collected on sugar maple only, but in 1989 all hardwood species were included. This report will focus on the crown rating system in relation to the QA aspects of the project. The tree crowns were rated for branch dieback, foliage transparency, foliage discoloration, and dwarfed foliage. Many similar types of measurements have been used in other countries to monitor tree crown condition (UN-ECE, 1988; Magasi, 1988). The crowns were rated using visual estimates for each crown condition rating. The same class codes were used for all four crown condition variables. Table 2 shows the percentage classes, the range of values in each class, and the acceptable range of deviation between remeasurements for all crown rating measurements. The 0 class was used to describe absolute zero for each particular crown condition rated.

Branch dieback was defined as branch mortality that began at the terminal portion of the branch and proceeded inward toward the trunk. The branch dieback measurement estimated the volume of crown foliage lost from the premature death of branches. When whole branches were dead in the upper exposed portion of the crown, they were also included in branch dieback unless signs, such as breakage, were present. Dead branches within the crown or in the lower crown were assumed to have died from suppression, and were not included in the totals. Large branches without small twigs (less than 2 cm diameter) were assumed to have died much earlier. These branches were not included because the branch dieback measurement was only for recent branch dieback.

The amount of light visible through the foliated portion of branches was used to estimate foliage transparency. Foliage transparency was intended as a repeatable means of estimating foliar loss and/or lack of foliage production. For the purposes of this study, the change in transparency between two measurements was analyzed for relationships to tree decline.

Foliage discoloration was based only on the foliated portion of the crown and did not include areas with dead or missing branches. Foliage was considered discolored when the overall appearance of a leaf was more yellow, red, or brown than green. When there was doubt concerning foliage color, it was rated green and not included in the rating. Declining trees tend to have off-green foliage before dieback occurs, but this was not measured because of the difficulty of standardization between individuals. Marginal leaf scorch and similar partial discoloration were not recorded unless more than half of a leaf was affected. Premature fall coloration of leaves has been associated with decline of sugar maple (Houston et al., 1990). However, rapid changes in fall coloration over time, and the regional differences in timing preclude its use as a response measurement.

The dwarfed foliage measurements were also based on the foliated portion of the crown (area of branch dieback was excluded from the total crown area). Foliage was considered dwarfed when leaf size was less than one half the size of leaves on healthy trees nearby. The dwarfed foliage condition should have been present on a branch as a whole (leaves near the branch tips normally tend to be smaller than the leaves further back). When in doubt, dwarfed foliage was not recorded.

Particular QA concerns in developing an appropriate QA program for this project relevant to data comparability included: 1) the use of multiple crews, 2) the large geographic region, and 3) the use of relatively untested visual estimates of tree crown conditions. Also, because the methodology was relatively new, it was extremely important to evaluate the QA information, including quality control (QC) data, in the context of project goals so that project management could make improvements.

To optimize data quality for the North American Sugar Maple Decline Project the following areas were addressed in project implementation:

1. The development of detailed documentation of the measurement systems, especially the crown rating measurements;

2. The use of annual training sessions for all participating crews;

3. The use of annual, independent audits to review field crew procedures;

4. A remeasurement of a subset of plots to assess the precision of the data;

5. The use of the remeasurement data, audit information, and training session evaluations to improve the project in the following years.

Methodology Documentation and Instruments

A comprehensive field manual with instructions for data collection was developed. The use of a standard manual is a key to ensuring data comparability (Garfield, 1984). Some methods were adapted from existing published procedures, i.e. dbh. In addition, a project QA plan was developed which described all activities and was designed to ensure and quantify data quality.

The crown condition ratings were used as the response measurements of tree vigor. The change in ratings was used as a measure of decline or recovery of tree vigor. The crown condition ratings were subjective measures that used visual estimates, and the use of visual estimates can create many interpretation problems (Innes, 1988). By necessity, these measurements were visual estimates and standardization was the greatest concern to ensure repeatability and comparability of measurements. Repeatability of these measurements was assured through development and field testing of standard procedures, intensive training, use of guides, and use of two persons to rate each tree.

The expected precision for the crown rating was one class deviation from the original rating. Changes in the crown condition rating needed to be greater than one class to be considered significant.

Two measurement tools were used in this study. The crown dieback grid (CDG) was patterned after a similar grid used by Doug Allen (SUNY), in his sugar maple studies; and the crown transparency standards (CTS) were produced based on suggestions made by Manfred Mielke (USDA Forest Service). Both aids were provided to all field crews to assist in the estimation of the proportions of crown damage and were included in the methods manual. The CDG was a standard square and dot grid commonly used to estimate areas on maps. The size of the area is not standardized, since it is intended for the determination of proportions. The basic approach is to draw an outline of the whole crown on a gridded plexiglass sheet and then draw in areas with branch dieback (or other crown rating measurement). After counting the number of grid points, the smaller number (affected area) is divided by the larger number (whole crown) and the answer is multiplied by 100 to determine the percentage of crown damaged.

The other aid was a CTS, used for foliage transparency measurements. The CTS was a series of rectangular grids or crown silhouettes that presented different proportions of black and white. The black represented the foliage and the white was the sky light. The class is indicated underneath the square, while the percentage range is given in the upper right corner. The basic principle for the use of CTS is to compare the tree crown with the CTS and then select the percentage class that most closely represents the tree crown.

Other practices were implemented to assist in the standardization of the crown ratings. A set of binoculars was used for closer examination of tree crowns. Each crown rating measurement was made by two persons from the opposite sides of a tree. When ratings differed, they exchanged positions and repeated the measurements. If the individuals still did not agree on a common rating, the average from the two observers was determined and recorded. This procedure decreased the amount of individual variability.

Training

All of the personnel establishing plots, taking measurements and making crown estimates attended a one-to-two day training session organized by the project staff. The field crew supervisors, QA auditor, and other project personnel (i.e., data management personnel) participated in the training session. The project coordinators (who also developed the procedures) conducted the training sessions. Annual retraining was required before data collection. Training was directed toward standardization of crown rating methodology, because the response measurements were qualitative, (i.e., visual estimates of crown characteristics) and do not have a finite or quantitative value.

The training session was structured to provide instructions on the interpretation and implementation of the crown rating methods. Training was designed for two-way communication so that crews could ask questions about particular situations or trees and was continued until crews were comfortable with the methodology, usually less than 1/2 day.

The second aspect of training was the performance testing phase. A test course of trees was developed and all individuals were tested and evaluated. After testing, a frequency distribution was developed for each tree from all participant results, and the distribution was used to determine the "actual" value (usually the median). Individuals with 10% of their values exceeding the allowable deviation (see Table 2) were provided with additional training prior to collecting field data.

Audits

The purpose of the audits was to review the implementation of the methodology. Audit reports were distributed to the appropriate national coordinator (U.S. or Canada) and state or provincial coordinator. Through the audits, discrepancies were discovered regarding method implementation. The major problem area was individual interpretations of statements and definitions in the methods manual. In the 1988 field season, this concern was noted for foliage discoloration and dwarfed foliage.

A pest outbreak of pear thrips (Taeniothrips inconsequens Uzel) in southern Vermont and western Massachusetts created unusual crown conditions. In the severely defoliated areas, the trees refoliated with a second growth flush during the late summer. In areas where the defoliation was moderate, the trees did not lose their leaves which created crowns of tattered leaves with brown edges. The foliage discoloration rating was supposed to ignore these leaves in the measurement but to include the condition in the notes. Unfortunately, various crews interpreted this crown condition in different ways. The situation was identified during the audits and highlights the necessity and advantage of audits.

The other use of the audits was to record problems the crews encountered with particular trees or situations. Not all potential conditions could be covered during the training sessions, so the audits helped to reduce interpretive difficulties. Crews also were encouraged to record any concerns or problems and to convey this information to their particular state or provincial coordinator.

Remeasurement System

Data quality assessments were made to estimate the precision of the crown ratings by a remeasurement program. Remeasurements for the other variables (i.e. dbh) were not taken. A target of five percent of the total plots was set for remeasurement. The percent of the trees remeasured was 5.39% and 8.32% in 1988 and 1989, respectively. The major focus was to quantify crew measurement precision for the crown rating

measurements. At the end of the field season the remeasurement information was compared to the original values to quantify the measurement precision.

Remeasured values were considered the same if they were $\pm$ 1 class from the original value for dominant and codominant trees or $\pm$ 2 classes from the original for intermediate and suppressed trees. Remeasurements were also evaluated depending on the nature of the original and remeasurement crews. A portion of the remeasurements were conducted by the original crew and estimated within-crew precision. Remeasurement data collected by a different crew estimated between-crew precision. It was expected that the within-crew precision would be greater than the between crew precision, and the analysis of the 1988 and 1989 data supported this theory (data not included).

The refoliating of crowns due to pear thrips defoliation created a situation where the leaf size changed rapidly during the measurement period. Some of the crews ignored the effect of refoliation on leaf size but other crews included this in the dwarfed foliage measurement. The 1988 remeasurement data for dwarfed foliage (Figure 1, top) illustrates this problem. The remeasurement information from 1988 was used to improve the method and training session for the following year (Figure 1, bottom). All of the figures are set up in the same manner. Each of the numbers represents the frequency of occurrences for that pair of original and remeasurement values. A blank designates that there were no observations at that particular pair.

Discussion

A demonstration of the success of this QA program can be seen in Table 3, where the measurement agreement percentages are shown for 1988 and 1989. In all four crown rating categories, the agreement percentage increased in 1989 from 1988 values. Major increases were noted in foliage discoloration and dwarfed foliage measurements, + 7.8% and + 10.5%, respectively. The improved agreement resulted from the elimination of the interpretative problems associated with the two crown condition variables and the absence of pear thrips damage in 1989.

The intensive crew testing and training of the crown condition measurement system resulted in 95.6 and 90.9 percent agreement in 1988 for branch dieback and foliage transparency, respectively (Table 3). This information is presented for 1988 and 1989 for branch dieback and foliage transparency, (Figures 2 and 3, respectively) and shows few outlying data points.

The between-crew measurement precision for 1988 and 1989 is shown in Table 4. All crown rating between-crew agreement percentages increased dramatically in 1989. This situation is most likely due to high crew retention from 1988, improved training sessions, and improved methodology.

The use of the audits was only one of many good communication channels developed in this project. The channels of communication included annual meetings to discuss problems or concerns, project management and local coordinators visiting field crews, and audits. All of these activities provided a simple and easy implementation of the project's objectives, especially the QA activities.

Conclusions

The results demonstrate that QA activities can improve and quantify the quality of the data for subjective measurements. In addition, the use of QA information and QC data can be applied to improve data quality in subsequent following years.

The following general procedures are recommended:

1. Develop and test consistent, well-documented procedures to ensure high data quality.

2. Implement training prior to data collection procedures to ensure crew familiarity with methodology.

3. Use audits for identifying problems concerning the implementation of methodology.

4. Build QA information into all programs so that data quality can be assessed and improved.

The research in this report has been funded by the U.S. Environmental Protection Agency and the USDA Forest Service through the Quality Assurance Project and Eastern Hardwoods Research Cooperative of the Forest Response Program of the National Acid Precipitation Assessment Program and Forestry Canada. This document has been prepared at the EPA Environmental Research Laboratory in Corvallis, OR, through contract # 68-C8-0006. It has been subjected to the Agency's peer review and administrative review and is approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Acknowledgements

The data for this document were provided by the cooperating groups in the "North American Sugar Maple Decline Project", (USDA Forest Service, Forest Pest Management; Forestry Canada; Maine Forest Service; University of Massachusetts; Michigan State University; New Hampshire Department of Resources and Economic Development; SUNY College of Environmental Science and Forestry; Vermont Agency of Environmental Conservation; and Wisconsin Department of Natural Resources). The authors would like to thank all the cooperating groups for providing the data used in this publication. In addition, appreciation is extended to the reviewers for providing constructive comments which improved the overall quality of the paper.

Literature Cited

Cline, S.P. and W.G. Burkman. 1989. The role of quality assurance in ecological research programs. In, Air Pollution and Forest Decline, J.B. Bucher and I. Bucher-Wallin, eds., Proc. 14th International Meeting for Specialists in Air Pollution Effects on Forest Ecosystems, IUFRO P2.05, Interlaken, Switzerland, Oct. 2-8, 1988, pp.361-365.

Garfield, F.M. 1984. Quality Assurance Principles for Analytical Laboratories. Assoc. of Official Analytical Chemists, Arlington, VA. 212 p.

Houston, D.R., D.C. Allen, and D. Lachance. 1990. Sugarbush management: a guide to maintaining tree health. USDA For. Serv. Gen. Tech. Rep., NE-129. 55 p.

Innes, J.L. 1988. Forest health surveys: problems in assessing observer objectivity. Can. J. For. Res. 18:560-565.

Lachance, D. 1985. Repartion geographique et intensite du deperissement de l'erable a sucre dans les erablieres au Quebec. Phytoprotection 66:83-90.

Magasi. L.P. 1988. Acid rain national early warning system: manual on plot establishment and monitoring. Forestry Canada, Ottawa, Ontario, Info. Rep. DPC-X-25.

Medlarz, S. 1988. The role of quality assurance in the Spruce-Fir Forest Research Cooperative, In, Proc. of the US/FRG Research Symposium: Effects of Atmospheric Pollutants on the Spruce-Fir Forests of the Eastern US and Federal Republic of Germany, Burlington, VT, Oct. 19-23, 1987. USDA For. Serv. Gen. Tech. Rep., NE-120, pp. 61-65.

Millers, I., D.S. Shriner, and D. Rizzo. 1989. History of hardwood decline in the Eastern United States. USDA For. Serv. Gen. Tech. Rep., NE-126. 75 p.

Millers, I., D. Lachance, and W.G. Burkman. 1990. North American sugar maple decline project - field methods. USDA For. Serv. Gen. Tech. Rep., (in preparation).

UN-ECE. 1988. Manual on methodologies and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests. Convention on Long-Range Transboundary Air Pollution, UN-ECE, Geneva, Switzerland.

Table 1. Distribution of clusters by political unit and the cooperating state or provincial agency.

Political Unit	Cooperating Agency	No. of Clusters	
United States:			
Maine	Maine Forest Service	18	
Massachusetts	Univ. of Massachusetts	10	
Michigan	Michigan State Univ.	10	
New Hampshire	NH Div. of Forest and Lands	6	
New York	State Univ. of NY, Syracuse	18	
Vermont	VT Dept. of Forests, Parks, and Recreation	26	
Wisconsin	WI Dept. of Natural Resources	18	
Subtotal			106
Canada:			
New Brunswick	Forestry Canada, FIDS*	10	
Nova Scotia	Forestry Canada, FIDS	2	
Ontario	Forestry Canada, FIDS	24	
Quebec	Forestry Canada, FIDS	24	
Subtotal			60
Total			166

* FIDS - Forest Insect and Disease Survey. Note these crews also included provincial personnel assistance.

Table 2. Percentage classes and acceptable range of errors for estimates of branch dieback, foliage transparency, foliage discoloration and dwarfed foliage.

Class Code	Percentage Class	Class Range	Acceptable Error Range
00	0	0	0 - 5
05	5	1 - 5	0 - 15
10	10	6 - 15	1 - 25
20	20	16 - 25	6 - 35
30	30	26 - 35	16 - 45
40	40	36 - 45	26 - 55
50	50	46 - 55	36 - 65
60	60	56 - 65	46 - 75
70	70	66 - 75	56 - 85
80	80	76 - 85	66 - 95
90	90	86 - 95	76 -100
99	100	96 -100	86 -100

Table 3. Measurement precision for sugar maple crown condition ratings for the 1988 and 1989 field seasons.

Crown Category	'88 % Agreement*	'89 % Agreement	Change ('89 - '88)
Branch Dieback	95.6	96.7	+1.1
Foliage Trans.	90.9	91.8	+0.9
Foliage Discolor.	91.3	99.1	+7.8
Dwarfed Foliage	88.3	98.8	+10.5

* % Agreement: ± 1 class between original and remeasurement data for dominant and codominant trees and ± 2 classes between original and remeasurement data for intermediate and suppressed trees.

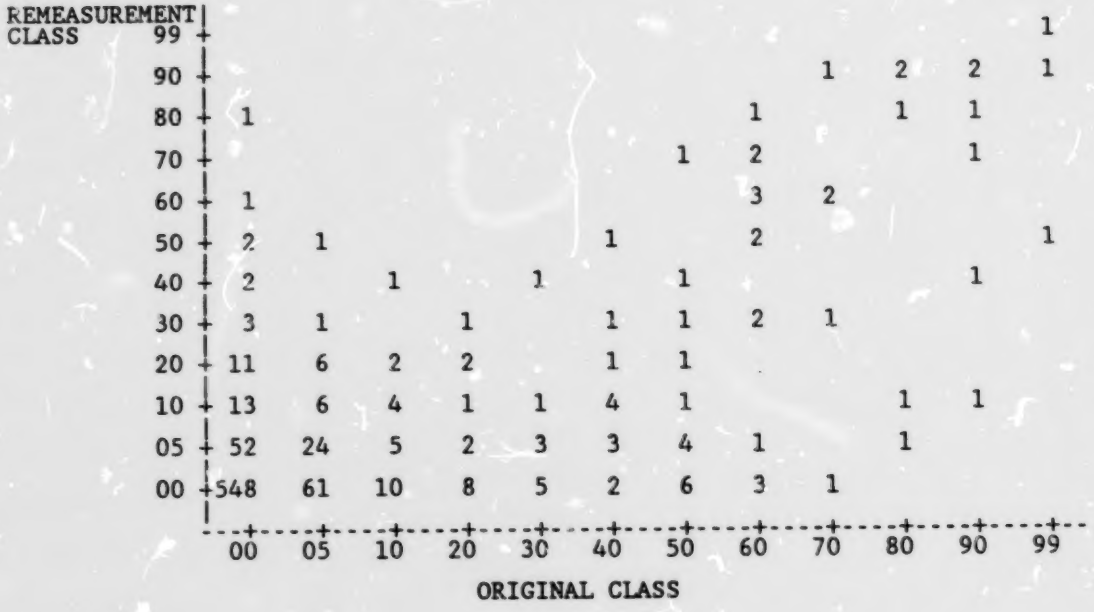
Table 4. Between-crew crown rating measurement precision for sugar maple, for the 1988 and 1989 field seasons.

Crown Category	'88 % Agreement*	'89 % Agreement	Change ('89 - '88)
Branch Dieback	79.2	95.6	+16.4
Foliage Trans.	74.3	93.8	+19.5
Foliage Discolor.	74.4	99.3	+24.9
Dwarfed Foliage	73.0	99.9	+26.9

* % Agreement: ± 1 class between original and remeasurement data for dominant and codominant trees and ± 2 classes between original and remeasurement data for intermediate and suppressed trees.

Figure 1. 1988 and 1989 original vs. remeasurement rating class for dwarfed foliage, sugar maple only (Note: Point is marked by number that designates frequency).

1988



1989

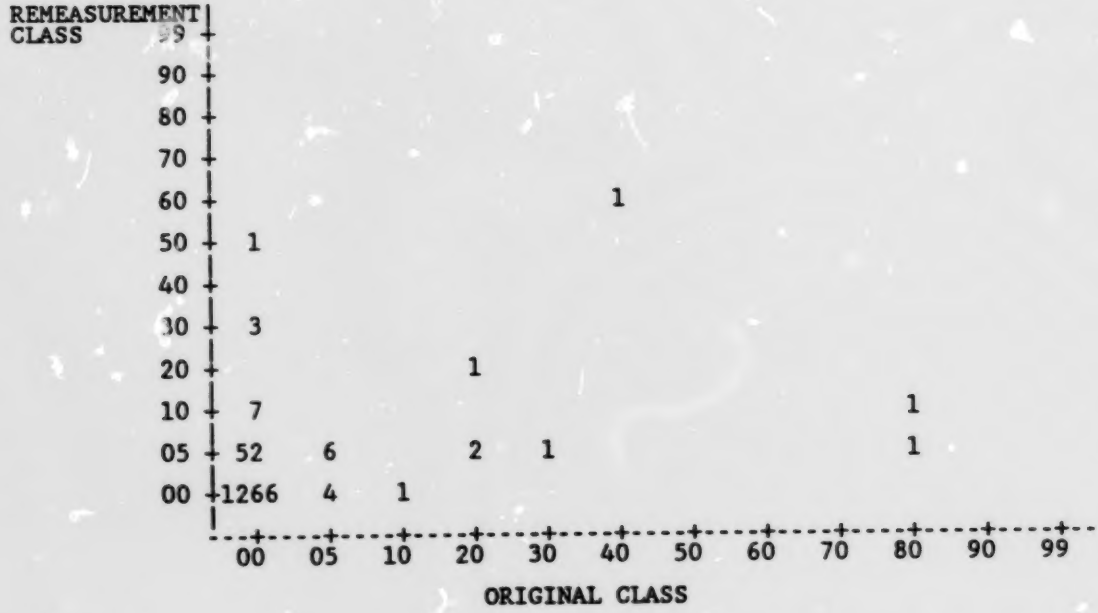
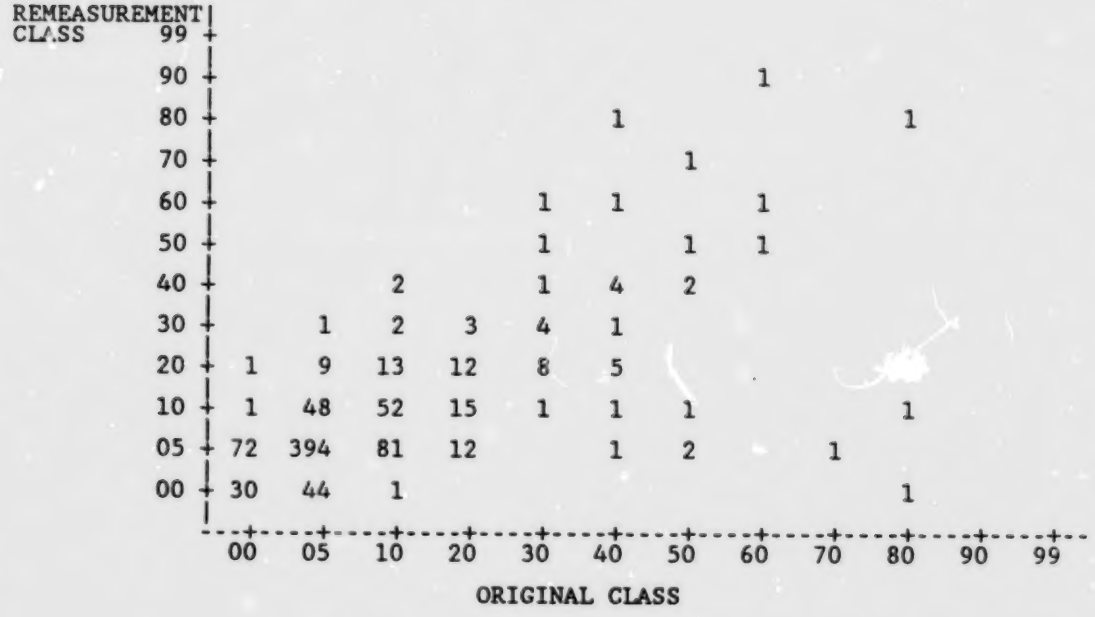


Figure 2. 1988 and 1989 original vs. remeasurement rating class for branch dieback, sugar maple only (Note: Point is marked by number that designates frequency).

1988



1989

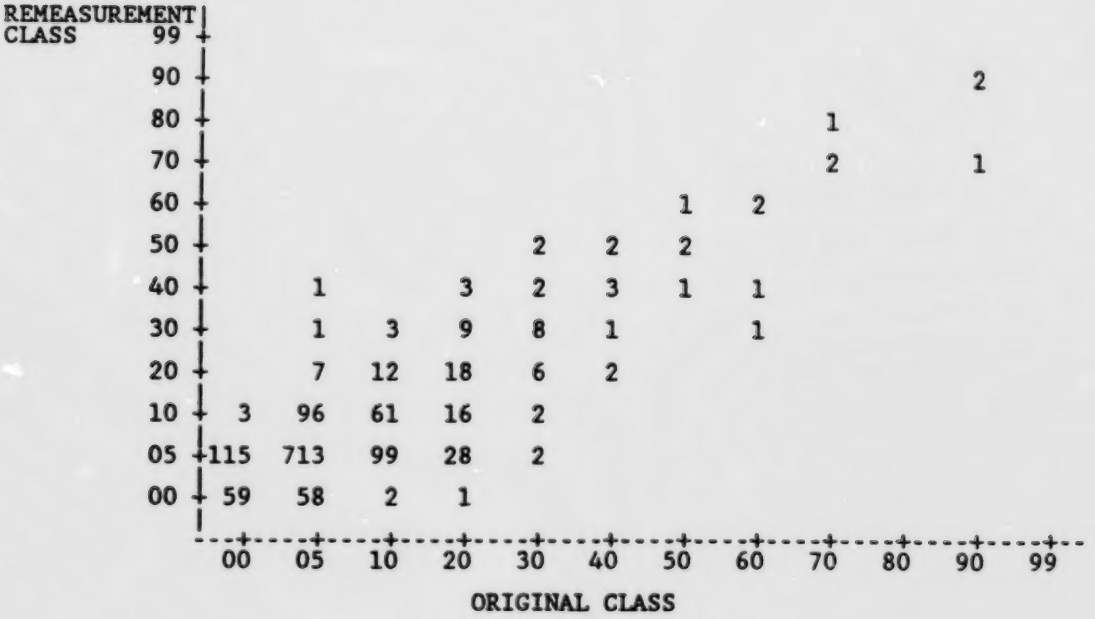
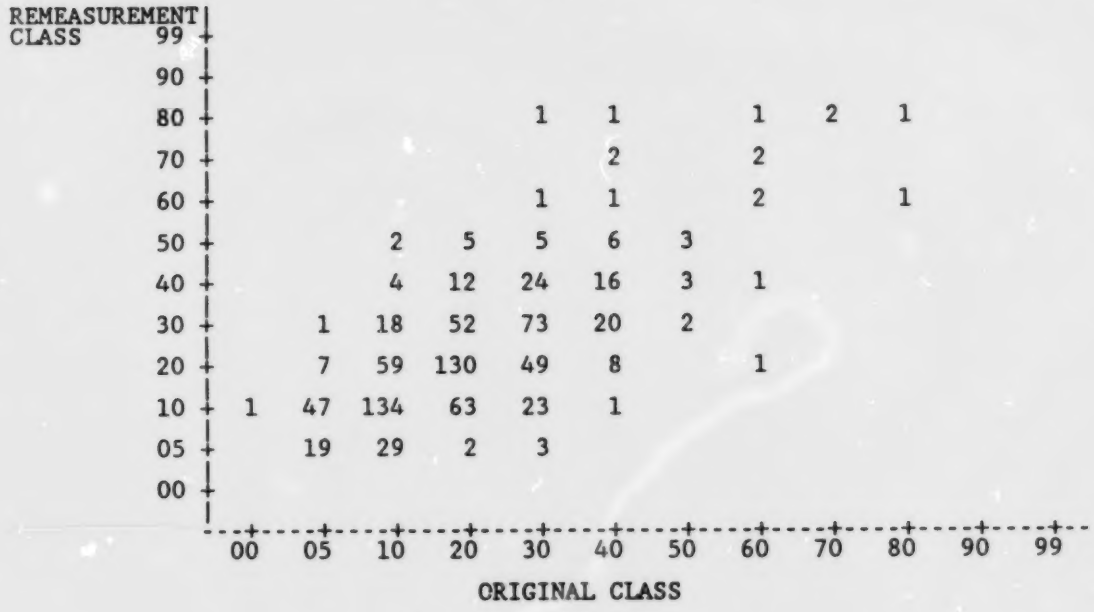
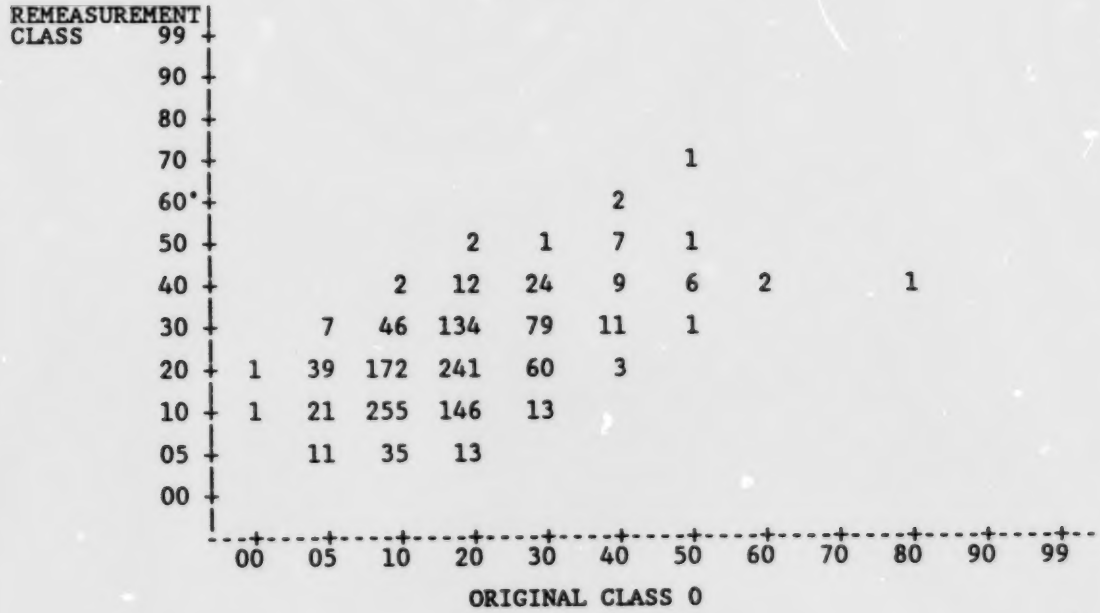


Figure 3. 1988 and 1989 original vs. remeasurement rating class for foliage transparency, sugar maple only (Note: Point is marked by number that designates frequency).

1988



1989



THE ROLE OF QUALITY ASSURANCE IN THE ASSESSMENT OF FOREST CONDITION IN EUROPE

John L. Innes
Forestry Commission
Alice Holt Lodge
Wrecclesham, Farnham
Surrey GU10 4LH
England

Abstract

There has been a major investment throughout Europe in the establishment of a monitoring programme of forest condition. The purpose is to collect accurate and representative data on changes in forest health in Europe. Although various protocols have been established, numerous problems exist over the quality management, quality assurance, quality control and quality evaluation of the programme. These problems severely limit its value, to the extent that the programme only partly fulfils its objectives and policy requirements. Despite these problems the current large-scale surveys are of value particularly if they are undertaken annually as part of a monitoring programme.

Introduction

In the late 1970s, it became apparent that forest decline either had occurred or was occurring in a number of areas in Europe. Such declines have occurred in the past and are well-documented, but the apparent extent and severity of the new problem led many people to believe that a major international problem was occurring. At the same time, there was a rise in interest in transboundary air pollution, particularly in the form of acidic deposition. A link between air pollution and forest decline was quickly postulated and, during the 1980s, many theories were put forward regarding the nature of the cause-effect relationships. Although many scientists believe that such a link exists, substantiation has proved to be surprisingly elusive (Prinz and Krause, 1989). Detailed research has not demonstrated a connection between forest decline and air pollution and, increasingly, research results are suggesting that factors other than pollution are more important. At the same time, many studies of increment have failed to demonstrate a regional or international growth decline, although it is clear that growth has declined in a number of localities.

Concern about forest decline led to pressure to monitor the condition of

forests and, since 1983, a steadily increasing number of European countries have commenced annual assessments of forest condition. To a certain extent, these have been formalized in a United Nations/Economic Commission for Europe protocol (UN-ECE, 1988), but it is apparent that major differences still exist between countries. Furthermore, it is becoming increasingly clear that the type of assessments that are currently the most favoured may not provide the answers to some of the key policy issues which the surveys are supposed to resolve. Many of the current problems relate to the lack of any formal quality assurance programme and, for example, there has been no attempt to introduce the type of programme run by the United States Environmental Protection Agency for the various environmental programmes in the U.S. (Mickler and Medlarz, 1987; Cline and Burkman, 1989). This paper addresses the following issues:

1) What are the primary objectives of monitoring forest condition? To what extent does the existing programme meet these objectives?

2) How subjective are the measurements? Can the observations made by different observers within a country be compared? To what extent can results from different countries be compared?

3) Can the natural year-to-year variation in forest condition be determined and separated from longer-term anthropogenic effects? How long a time series will be required before a forest "decline" can be recognised? Is there any evidence of a European-wide decline of forests?

4) Do large-scale representative surveys of tree condition provide information on the effects of air pollution on forests?

Objectives of the Monitoring Programme

The stated objective of the United Nations International Cooperative Programme on Assessment and Monitoring of Air Pollution Effects on Forests is "to facilitate the collection, on a national level, of comprehensive and comparable data on changes in forests related to actual environmental conditions (in particular air pollution, including acidifying depositions) as a basis for, e.g., improved evaluation of trends in damage resulting from pollution and establishment of a better understanding of dose/response relationships" (UN-ECE, 1988, p. 4). The programme is divided into three main parts, dealing with large-scale representative surveys of forest health, intensive studies on permanent plots and special forest ecosystem analyses. To date, most of the work has concentrated on the large-scale surveys, and this paper will focus on these.

The specific objective listed by UN-ECE (1988) for the first part of the programme is "the periodic assessment of changes in the health status of forests by collecting representative data on the nature, extent, and intensity of forest damage" (UN-ECE, 1988, p. 9). To do this, a systematic sampling design has been adopted by

all countries in Europe. Sample plots are located on a 16 x 16 km grid, with plots being located at those grid intersections which occur in forest (defined as an area of woodland that is sufficiently large to enable the sampling to take place, i.e. 1 ha or greater). The same programme has been adopted by the European Economic Community (EC) which has passed legislation requiring Member States to undertake annual representative surveys. This legislation states that plots should only be located in forest tree stands with at least 20% canopy closure at rotation age, with minimum plot sizes of 0.5 ha.

A systematic sampling design is not necessarily the best suited to the problem (Innes, 1988a). Where the forest within a country is distributed non-randomly, systematic sampling may result in gaps in the coverage. For example, in Great Britain, forests are mostly located in Wales and Scotland, but the highest levels of pollution occur in east-central England and the systematic sampling design resulted in no plots being located in the most polluted areas. There appears to have been little discussion at the outset of the international programme as to the most efficient way in which suitable data might be collected and, in retrospect, it is apparent that different sampling designs and sampling densities might have been more appropriate in different countries.

At each plot location, a number of trees are selected for monitoring. The number varies but the generally accepted practice is to locate six trees in each of four sub-plots within the stand. The sub-plots are usually arranged in a cross-cluster, with each arm of the cross being 25 m long. However, in very dense stands, it may be necessary to extend the lengths of the arms in order to locate the sub-plots at positions where the crowns of the trees are visible. Frequently, this means that the trees sampled in dense plantations are on the edges of stands. The result is that the standard cross-cluster is often restricted to older or semi-natural woodland, whereas the extended cluster is used for younger plantations. No attempt has been made to examine the statistical effects of this potential bias.

In both the UN-ECE and EC monitoring programmes, the two main criteria used to characterise the health of trees are crown density and crown discoloration. Greater emphasis is placed on crown density, which is usually interpreted as the proportion of leaves or needles that have been lost compared with some 'ideal'. However, it is assessed by comparison with a reference tree(s) and is actually a measure of crown transparency rather than the number of needles or leaves lost (Mahrer, 1989). For example, Innes and Boswell (1990) have shown that there is a poor relationship between crown density and the number of needle-years retained by Sitka spruce (Picea sitchensis (Bong.) Carr.), Norway spruce (Picea abies (L.) Karst.) and Scots pine (Pinus sylvestris L.).

The significance of crown density as a measure of tree health is unclear. Attempts to relate crown density to increment have produced variable results (Innes and Cook, 1989), because, obviously, various factors other than crown density affect increment. There is no universal threshold at which a reduction in crown density begins to affect increment, with studies indicating that the crown density - increment relationship may

be site-specific (Schmid-Haas, 1989) or even dependent on the individual tree (Schlaepfer, 1986). Consequently, there is considerable doubt over the value of crown density as a measure of tree health.

Crown density scores are normally reported in a number of unequally sized categories. These relate to losses in density of 0-10%, 11-25%, 26-60%, 61-99%, and dead. The system is based on that adopted by the Federal Republic of Germany in 1983 (Anon, 1984). Since then, it has been widely accepted that the classes may be too broad to allow subtle changes in forest condition to be identified (Innes and Boswell, 1990). Several countries used different systems (e.g. 20% classes in Norway, 10% classes in Great Britain and 5% classes in Switzerland) and, in 1989, the EC revised its legislation to make its Member States use 5% classes. However, while assessments are to be made in 5% classes, there is a considerable amount of resistance to dropping the original reporting system because of the difficulty of comparing new results with those obtained using the original categories. Interestingly, the resistance has come at the administrative level rather than from the scientists and foresters involved in making the assessments.

The initial categorization was based on the belief that trees with more than 60% loss of foliage were incapable of recovery. Over the last few years, it has become apparent that this is not the case. For example, in Britain, a substantial proportion of the trees that were scored with density losses in excess of 60% in 1987 or 1988 had density losses of less than 60% by 1989 (Innes and Boswell, 1990). Similar results have been found in Switzerland where the progression of individual trees has been followed since 1985 (Hagi, 1989; Mahrer, 1989).

A number of alternative indices of tree health now exist (e.g. Alexander and Carlson, 1988; Lesinski and Westman, 1987; Gruber, 1989; Niehaus, 1989; Roloff, 1989; Westman, 1989; Westman and Lesinski, 1986) but most countries have not taken these up. Many of these, such as the extent of crown dieback and the frequency of dead shoots within the live crown, are likely to be better indicators of crown condition than crown density as they indicate the presence of a potential problem. In addition, it is possible to look at discoloration in more detail. The standard procedure involves the assessment of the proportion of discolored foliage into the same five categories as crown density. There is some confusion here: a number of countries only include yellowing, others include all forms of discoloration. In Great Britain, discoloration has been divided into yellowing and browning, and the proportions affected on conifers are divided into first-year needles and older needles. This is much more useful as a diagnostic tool and also facilitates the recognition of discoloration caused by autumnal senescence in Scots pine.

The analysis of such data is not easy as the variety of data collected is very difficult to interpret. Principal components analysis has been used to try and identify a composite index of tree condition (Innes and Boswell, 1988; 1989) but this has suggested that many of the indices are relatively independent of each other. Currently,

other statistical and mathematical techniques, such as correspondence analysis and two-way indicator species analysis, are being used to examine the data, but no results are yet available.

There are clearly considerable problems in using the existing surveys to meet the specified objectives of the programme. The indices that are assessed in most countries are poor indicators of tree health and their interpretation is difficult. They are not specific to the effects of air pollution and may even have little value for assessing whether or not a tree has been stressed. Better indicators exist, but there has been a considerable amount of resistance to adopting these.

Subjectivity of Assessments and Comparisons Between Different Countries

Assessments of crown density are made by comparison with reference photographs (e.g. Bosshard, 1986; Gellini and Clauser, 1986) or reference trees within the immediate vicinity or region (Landmann, 1989). Investigation of the training needs and comparisons between observers indicates that there is an initial learning curve after which observations stabilize (Innes, 1988b). Approximately one week of training is required and this needs to be repeated each year, even if the same observers are used.

Variations between observers can be reduced by adopting strict criteria for the assessment of crown density. One of the most important causes of variation is differences in the amount of crown included in the assessment. This is particularly important for plantation trees, where shading may result in branch death in the lower crown. A definite rule needs to be adopted about such branches that can be universally applied. So far, there appears to have been no international agreement about this.

Checking the variation between observers is one of the most important roles for quality control in the monitoring programme. There are several ways that this can be done, but usually, one or more teams of observers visit all survey teams during the field season. The use of pairs of observers can reduce individual bias but, frequently, the observations of one of the individuals tend to predominate. The extent of bias can be determined by the average difference between the scores obtained by the survey team and the check team. In addition, it is also useful to determine the degree of variation.

In Great Britain, bias is evaluated on the basis of a minimum sample of 24 trees of each species for each team. To date, systematic bias has been within 1%, which has been arbitrarily considered as acceptable, and 95% of the observations (two standard deviations about the mean) have generally been within the range $\pm$15%, which again is considered to be acceptable (Innes and Boswell, 1988; 1989). These results are very similar to those obtained during checks of observers in Switzerland (Mahrer, 1989). It is important to bear in mind this variation when examining the progression of individual trees through time.

Bias through time is more difficult to evaluate. There is a definite possibility of observer drift, particularly for the control team. It is difficult to see how this can be avoided, although it should be less of a problem for those countries using photographic standards than for those using local or regional standards.

The comparison of the condition of trees in different countries was one of the primary aims of both the UN-ECE and EC programmes. However, lack of standardization of detailed methodologies (Innes, 1988a) makes comparisons difficult. This is now recognised by both the UN-ECE (UN-ECE, 1990) and the EC (Commission of the European Communities, 1989). Nevertheless, "league tables" of forest condition are still published by international agencies, frequently without any of the caveats that would enable their proper interpretation. The lack of explanatory information in some reports has led to serious misinterpretation of the results (Innes, 1988c), with major implications for policy decisions at various levels.

Identification of Trends Through Time

The term "forest decline" implies that there has been a decline in forest condition. While this is clearly the case for some localities, it is much less obvious on a national or international scale. Most surveys started in the period 1984-86 and the longest reliable time series that is available is therefore only 6 years. An example of such a series, for the Federal republic of Germany (Anon, 1989), is shown in Figure 1. The proportion of trees with more than 60% reduction in density has remained constant over the period and there have only been minor changes in the proportion of trees with more than 25% reduction in density. Overall figures such as these obviously may conceal interspecific differences or regional trends, but they clearly suggest that arguments for a major decline in tree health during the 1980s cannot be supported.

Time series such as that shown in Figure 1, rely on the consistency of data from year to year. In some cases, their reliability has been overestimated. For example, there may have been changes in the perception of the observers over the period. Alternatively, assessment techniques may have been subtly altered, resulting in inconsistent series. In Great Britain, the first surveys were undertaken in 1984. However, observer training was very limited and inadequate standards were set. With the introduction of photographic standards in 1987, there appeared to be a sharp deterioration in forest condition, as judged by the overall statistics (Innes and Boswell, 1987). However, individual survey teams were adamant that they had not seen a sharp deterioration in the condition of the sample trees. Consequently, 1987 was taken as the base year for the time series, leading to accusations that the "goal posts" had been changed. A further modification occurred in 1989 when the EC introduced the 5% classification system. Previously, trees had been assessed on a 10% scale and the switch has created a minor change in the manner in which the observers assess the trees. However, it has been possible to take this into account when assessing the overall statistics and three years of data are therefore available. The data are presented in Table 1, from which it can be seen that the annual changes in condition have been

relatively small.

There is very little information on the natural year-to-year variation in crown density and this places severe limitations on the use of traditional methods of quality control (Innes and Boswell, 1989). Furthermore, there is no base-line against which changes can be measured. As a number of factors affect crown density, it seems likely that considerable inter-annual variation will occur. Episodic events, such as severe defoliation by insects and extreme climatic conditions, are likely to have a major effect, and these may persist for some time. Consequently, it will be very difficult to identify confidently trends in crown density over periods of less than ten years. For species that retain their needles for long periods, such as Sitka and Norway spruce, recovery periods may be even greater, with consequent effects on the length of time required to identify trends.

The Role of Air Pollution

Assessments of forest condition provide a description of the state of the crown of a sample of trees. Unless other information is gathered, very little interpretation can be made of such data. The EC surveys collect supplementary information in the form of presence-absence of easily identifiable causes of damage (abiotic agents, insects, fungi, game and grazing, action of man, fire, known pollution, other). Within the EC, 27% of trees in 1987 and 13% in 1988 were recorded as being damaged by a known pollutant (Commission of the European Communities, 1989). However, it is not clear how the damage was identified. Given the non-specificity of chronic pollution symptoms, it seems likely that the figures refer to either acute damage by local pollution or the damage has been misidentified. As acute damage is relatively rare, the latter explanation is more likely.

An alternative approach is to examine spatial trends in tree condition to see whether there is any correlation with pollution. There have been few attempts to do this although interest appears to be developing in the approach (e.g. Neumann, 1989; Stock, 1988; Innes and Boswell, 1988; 1989; Landmann and Fabbri, 1989). Early attempts (e.g. Binns et al., 1985; 1986; Innes et al., 1986) were based on multiple linear regression, using a limited number of explanatory variables. However, there are clearly a large number of variables that could potentially influence tree condition and these need to be taken into account. Many of these variables are highly intercorrelated (Innes and Boswell, 1988; Neumann, 1989) creating major problems for any regression analysis. This was confirmed by Innes and Boswell (1988), who used ridge regression to demonstrate the instability of the multiple regression coefficients. A further difficulty that has been encountered is that many of the significant correlations that can be identified between indices of crown condition and environmental variables are caused by a small number of extreme values (Innes and Boswell, 1989).

Table 1. Comparison of tree crown density results for the period 1987-1989. Class 0 represents 0-10 per cent reduction in density, class 1 represents 11-20 per cent reduction, class 2 represents 21-30 per cent reduction and so on. The numbers of trees compared within each species are given. The figures in the columns refer to the percentage of trees in each category.

		Crown density class									
		0	1	2	3	4	5	6	7	8	9
Sitka spruce	1987	14	23	25	19	12	5	2	0	0	0
n = 1223	1988	8	22	26	23	14	6	1	0	0	0
	1989	13	22	24	16	10	5	3	2	2	0
Norway spruce	1987	21	24	24	15	9	4	2	1	0	0
n = 1512	1988	21	27	24	16	7	4	1	0	0	0
	1989	27	28	22	14	6	2	1	0	0	0
Scots pine	1987	19	23	25	18	8	3	1	1	1	1
n = 1549	1988	8	23	32	20	9	4	1	1	1	1
	1989	13	31	27	15	6	3	2	1	1	1
Oak	1987	9	15	21	29	15	6	4	1	0	0
n = 695	1988	4	16	29	27	14	6	2	1	1	0
	1989	7	22	31	26	10	3	1	0	0	0
Beech	1987	8	22	26	28	12	3	1	0	0	0
n = 672	1988	9	24	33	22	9	2	1	0	0	0
	1989	19	31	29	16	4	1	0	0	0	0

Figure 1: Changes in the proportion of trees in the Federal Republic of Germany above a given level of crown density reduction, 1984-1989. (From Anon, 1989).

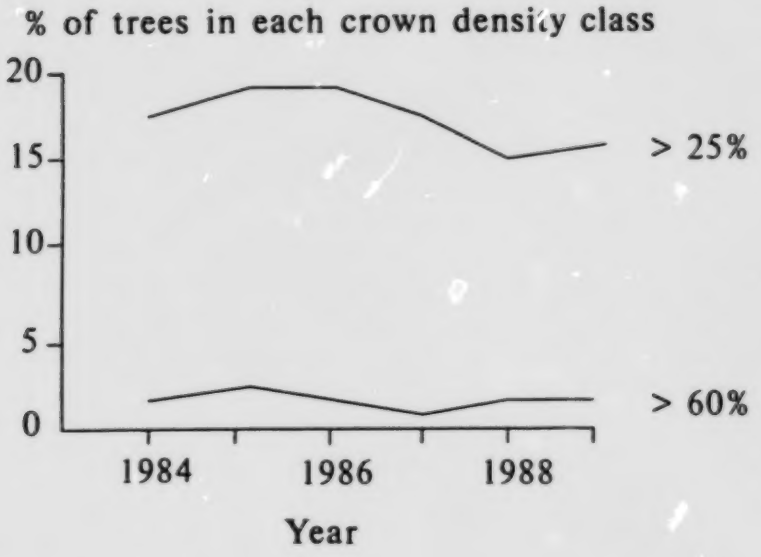
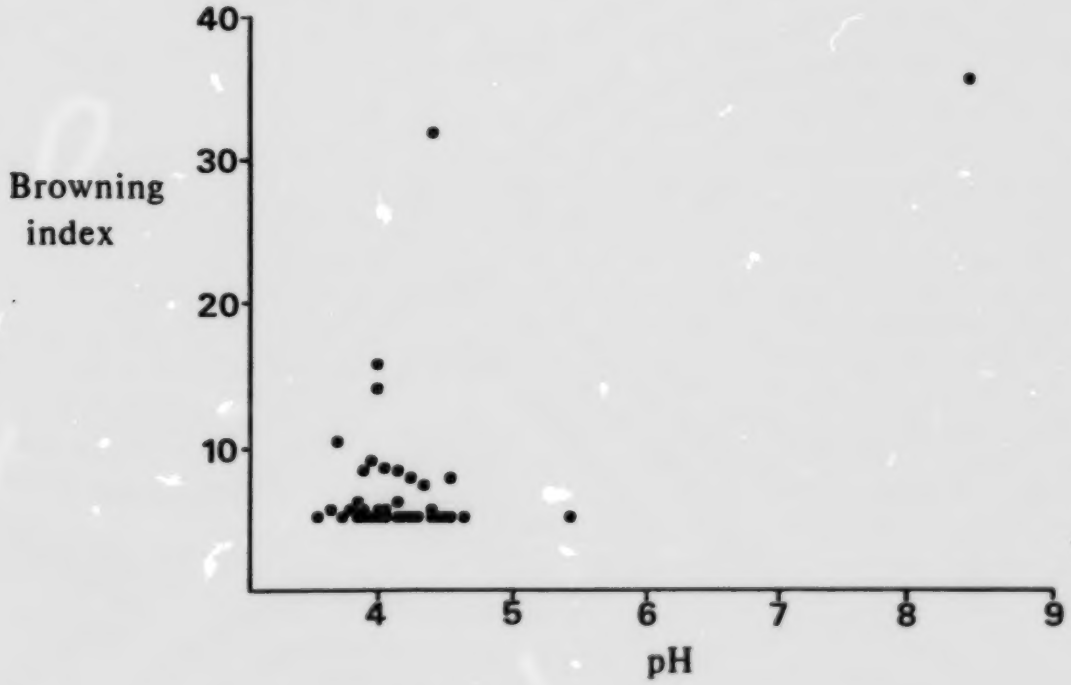


Figure 2: Relationship between the degree of browning of older Sitka spruce needles and soil pH at 10 cm depth. Each point refers to the mean value for a site.



An example of a spurious correlation can be illustrated by reference to data collected in Great Britain in 1988. A significant ($p < 0.001$) correlation was found between the extent of browning of older needles and the pH of the soil at 10 cm depth. However, the correlation is the result of a single out-lying value (Figure 2).

Apart from the statistical difficulties, there are problems with the actual data. The explanatory variables are rarely measured on site and, if so, there may be difficulties with their characterisation. For example, it is very difficult to ascertain the actual dose of specific pollutants received by the trees, yet this is crucial to any cause-effect explanation. Interactions between explanatory variables are difficult to take into account and it is likely that different combinations of stresses may produce the same end-result for the trees.

In view of the above difficulties, it may be more appropriate to use visual techniques alone. Indices of crown condition can be mapped and compared with maps of specific pollutants. Areas of interest can quickly be identified and supplementary studies to establish cause-effect initiated. Even this approach is problematical as it is apparent that crown condition is spatially highly variable. This creates difficulties for mapping, and some form of geostatistical analysis may be required before reliable maps can be prepared (Innes and Boswell, 1989). The technique also relies on the accuracy of any maps of the explanatory variables and it is important to bear in mind the limitations of using the best-available maps.

The studies that have been done have not demonstrated any clear link between air pollution and tree health. All, however, suggest that site quality may have an important effect on tree condition. This is hardly surprising as site quality has long been known to one of the primary determinants of forest productivity.

Conclusions.

The existing programme of large-scale representative surveys only partly achieves its stated objectives for a number of reasons. The surveys do not provide "representative data on the nature, extent, and intensity of forest damage". Instead, they provide limited information on two inadequate parameters of forest condition, crown density and crown discoloration. The data are, in some cases, representative; in others, the data are of little value in establishing the condition of forests in a region or country because of the poor coverage. Alternative sampling designs are available (Innes, 1988a), but it is unlikely that many countries will implement them. Similarly, more appropriate indices of tree health have now been identified (such as the extent of shoot death and dieback), but there is resistance over their inclusion into the annual surveys.

The information that is collected does not, in itself, reveal anything about the impact of transboundary air pollution on forests. However, it is frequently assumed to do so and there has been a clear slant in the summaries of some of the reports that have been published. These aspects can be analysed if supplementary information is

available, but there are many pitfalls and interpretations should be made with extreme care. Even when an association between an index of tree health and an environmental variable is identified, it will be necessary to undertake more detailed investigations to establish actual cause and effect.

Despite the above difficulties, large-scale surveys can be of value, particularly if they are repeated annually in the form of a monitoring programme. Annual changes in the condition of trees can be determined and related to specific environmental conditions during that year. If supplementary observations are made, such as radial and height increment, then it will be possible to determine the effects of silvicultural treatments and other factors on forest condition. Furthermore, the continued deterioration of forest stands in some areas is of concern and it is important to ensure that such deteriorations do not intensify or spread. Monitoring provides an ideal method of recognizing and quantifying the problem.

References

Alexander, S.A. and Carlson, J.A. 1988. Visual Damage Survey Pilot Test Project Manual. Blacksburg, Virginia Polytechnic Institute and State University.

Anon. 1984. Waldschadenserhebung. Bonn, Bundesministerium für Ernährung, Landwirtschaft und Forsten.

Anon. 1989. Waldzustandsbericht. Ergebnisse der Walschadenserhebung 1989. Bonn, Bundesministerium für Ernährung, Landwirtschaft und Forsten.

Binns, W.O., Redfern, D.B., Boswell, R.C. and Betts, A.J.A. 1986 Forest health and air pollution: 1985 survey. Forestry Commission Research and Development Paper 147. Edinburgh, Forestry Commission.

Binns, W.O., Redfern, D.B., Rennolls, K. and Betts, A.J.A. 1985. Forest health and air pollution: 1984 survey. Forestry Commission Research and Development Paper 142. Edinburgh, Forestry Commission.

Bosshard, W. (ed.) 1986. Kronenbilder. Birmensdorf, Eidgenössische Anstalt für das forstliche Versuchswesen.

Cline, S.P. and Burkman, W.G. 1989. The role of quality assurance in ecological research programs. In: Bucher, J.B. and Bucher-Wallin, I. (eds.) Air Pollution and Forest Decline. Birmensdorf, Eidgenüssische Anstalt für das forstliche Versuchswesen, 361-365.

Commission of the European Communities. 1989. European Community Forest Health Report 1987-1988. Luxembourg, Office for Official Publications of the European Communities.

Gellini, R. and Clauser, F. 1986. Prime indagini sul deperimento dei boschi in Italia. Ministero Agricoltura e Foreste, Roma. Collana Verde 72, 1-91.

Gruber, F. 1989 Phänotypen der Fichte (Picea abies (L.) Karst.). I. Verzweigungsphänotypen: Genotyp und Modifikation. Allgemeine Forst- und Jagdzeitung 160, 157-166.

Hägi, K. 1989. Terrestrische Waldschadeninventur. Berichte 314. Birmensdorf, Eidgenössische Anstalt für das forstliche Versuchswesen.

Innes, J.L. 1988a. Forest health surveys - a critique. Environmental Pollution 54, 1-15.

Innes, J.L. 1988b. Forest health surveys: problems in assessing observer objectivity. Canadian Journal of Forest Research 18, 560-565.

Innes, J.L. 1988c. Forest decline in urban areas - a comment. Science of the Total Environment 72, 227-230.

Innes, J.L. and Boswell, R.C. 1987. Forest Health Surveys 1987. Part 1: Results. Forestry Commission Bulletin 74. London, HMSO.

Innes, J.L. and Boswell, R.C. 1988. Forest Health Surveys 1988. Part 2: Analysis and Interpretation. Forestry Commission Bulletin 79. London, HMSO.

Innes, J.L. and Boswell, R.C. 1989. Monitoring of Forest Condition in the United Kingdom 1988. Forestry Commission Bulletin 88. London, HMSO.

Innes, J.L. and Boswell, R.C. 1990. Reliability, presentation and relationships amongst data from inventories of forest condition. Canadian Journal of Forest Research 20, in press.

Innes, J.L., Boswell, R., Binns, W.O. and Redfern, D.B. 1986. Forest Health and Air Pollution: 1986 Survey. Forestry Commission Research and Development Paper 150. Edinburgh, Forestry Commission.

Innes, J.L. and Cook, E.D. 1989. Tree-ring analysis as an aid to evaluating the effects of pollution on tree growth. Canadian Journal of Forest Research 19, 1174-1189.

Landmann, G. 1989. Observation au sol de l'état des forêts: évaluation critique des données 1983-1988. Journées de Travail DEFORPA. Nancy, INRA, 1.1.1-1.1.16.

Landmann, G. and Fabbri, B. 1989. Relations entre l'aspect des cimes, la croissance, les paramètres sylvicoles et stationnels. Le cas du sapin dans les Vosges. Journées de Travail DEFORPA. Nancy, INRA, 3.6.1-3.6.17.

Lesinski, J.A. and Westman, L. 1987. Crown injury types in Norway spruce and their applicability for forest inventory. In Perry, R., Harrison, R.M., Bell, J.N.B. and Lester, J.N. (eds.) Acid Rain: Scientific and Technical Advances. London, Selper, 657-662.

Mahrer, F. 1989. Problems in the determination and interpretation of needle and leaf loss. In Bucher, J.B. and Bucher-Wallin, I. (eds.) Air Pollution and Forest Decline. Birmensdorf, Eidgenössische Anstalt für das forstliche Versuchswesen, 229-231.

Mickler, R.A. and Medlarz, S.A. 1987. The role of quality assurance in national acid rain research in the United States. Environmental Technology Letters 8, 459-466.

Neumann, M. 1989. Einfluss von Standortsfaktoren auf den Kronenzustand. In Bucher, J.B. and Bucher-Wallin, I. (eds.) Air Pollution and Forest Decline. Birmensdorf, Eidgenössische Anstalt für das forstliche Versuchswesen, 209-214.

Niehaus, G. 1989 Terrestrische Zustandserfassung von Kiefernkronen (Pinus sylvestris) - ein differenzierter Anspracheschlüssel. Der Forst und Holzwirt 44, 200-206.

Prinz, B. and Krause, G.H.M. 1989. State of scientific discussion about the causes of the novel forest decline in the Federal Republic of Germany and surrounding countries. In Bucher, J.B. and Bucher-Wallin, I. (eds.) Air Pollution and Forest Decline. Birmensdorf, Eidgenössische Anstalt für das forstliche Versuchswesen, 27-34.

Roloff, A. 1989. Kronenentwicklung und Vitalitätsbeurteilung ausgewählter Baumarten der gemässigten Breiten. Schriften aus der Forstlichen Fakultät der Universität Göttingen 93.

Schlaepfer, R. 1986. Relationship between diameter growth and crown parameters, in particular foliage loss, of Picea abies and Abies alba: results of a pilot study in Switzerland. In: Fujimori, T., and Whitehead, D. (eds.) Crown and Canopy Structure in Relation to Productivity. Ibaraki, Forestry and Forest Products Research Institute, 70-92.

Schmid-Haas, P. 1989. Do the observed needle losses reduce increments? In Bucher, J.B. and Bucher-Wallin, I. (eds.) Air Pollution and Forest Decline. Birmensdorf, Eidgenössische Anstalt für das forstliche Versuchswesen, 271-275.

Stock, R. 1988. Aspekte der regionalen Verbreitung "Neuartige Waldschaden" an Fichte im Harz. Der Forst und Holz 43, 283-286.

UN-ECE. 1988. Manual on Methodologies and Criteria for Harmonized Sampling, Assessment, Monitoring and Analysis of the Effects of Air Pollution on Forests.

Geneva, Convention on Long-Range Transboundary Air Pollution.

Westman, L. 1989. A new method for assessment of visible damage to birch and other deciduous trees. In Bucher, J.B. and Bucher-Wallin, I. (eds.) Air Pollution and Forest Decline. Birmensdorf, Eidgenössische Anstalt für das forstliche Versuchswesen, 223-228.

Westman, L. and Lesinski, J.A. 1986. Kronutglesning och andra förändringar i grankronan. Naturvaordsverket Rapport 3262.

QA ASPECTS OF THE EMAP INLAND AQUATIC PROGRAM

Dan McKenzie
Environmental Research Laboratory
U.S. Environmental Protection Agency
Corvallis, OR 97333

Abstract

The Environmental Protection Agency is initiating an ecological monitoring program to monitor the condition of the nation's ecological resources. The program, titled Environmental Monitoring and Assessment Program, EMAP, includes components to address all the resource systems, lakes and streams, wetlands, forests, arid lands, agroecosystems, near coastal estuaries and is designed to monitor indicators of stressors, exposure and response. EMAP is based on a statistical design that will provide estimates of indicator status for populations of resources in regions of the country. These estimates when combined with additional information on landscape characteristics, stressor and exposure indicators, and ecological reference values will permit assessments of ecological condition and diagnostic investigations to evaluate possible causes of environmental condition.

This presentation addresses requirements for information quality and quality assurance within the Inland Aquatic components of EMAP, presented from the author's perspective as an EMAP program manager. A schematic model is presented and developed as a framework for the planning process to total quality management, a top-down focus, and for the development of DQO's for the data generation process, a bottom-up process. Two examples, mercury concentrations in fish and an index of biotic integrity (IBI), are presented to illustrate the process and identify the quality assurance issues that need to be addressed.

EMAP MANAGEMENT'S COMMITMENT TO TQM

- Committed to TQM Concepts
- Team Lacks Common Understanding
- Shared Frustration Over Traditional QA
- Challenged and Committed to do Better

EMAP AREAS FOR IMPROVEMENT

- Quality of Planning, Management, Commitment
- Quality of Design and Analyses
- Quality of Implementation
- Quality of Performance
- Quality of Actions

Quality of Planning, Management, Commitment

- Comprehensive Plans
- Meaningful Schedules
- Follow-through on Commitments

Quality of Design and Analyses

- Efficient Robust Design
- Ease of Use
- Meaningful Analysis, Summaries, Displays
- Responsive Information Management

Quality of Actions

- Reach or Produce Objectives and Goals
- Successful Balance Between Competing Interests
- Recognized by Peers, Opponents and Proponents
- Mission and Goals of Agency Supported

Quality of Implementation

- Meets or Exceeds DQO's
- Representativeness
- Detection Ability
- Optimal Allocation of Resources

Quality of Performance

- Results Available on Time
- Lives up to Expectations
- Understandable Communication
- Appropriate Scale Reported

EMAP Example:
Status and Trends in Hg Concentration in Fish in Upper-Mid-Western Seepage Lakes

- Historical QA/QC Focus - Hg Conc in Fish Tissue

EMAP Challenge:

- Identify Options, Uncertainties and Consequences
- Statistical Design and Analyses
- Sampling Methodology
- Taxonomy, Presence/Absence of Fish
- Age, Weight, Length Observations
- Hg Concentration

Comparison of 1990 Through 2005 Information Identifies Meaningful Changes

QUALITY ASSURANCE ASPECTS OF EPA'S NEAR COASTAL ECOLOGICAL STATUS AND TRENDS PROGRAM

J. F. Paul*
U.S. Environmental Protection Agency
Narragansett, Rhode Island 02882

Abstract

The U.S. Environmental Protection Agency is initiating an Environmental Monitoring and Assessment Program (EMAP) to monitor the status and trends of the nation's near coastal waters, forests, wetlands, surface waters, agroecosystems, deserts, and grasslands. This program is also intended to evaluate the effectiveness of Agency policies at protecting ecological resources occurring in these systems. Monitoring data collected for all ecosystems will be integrated for national status and trends assessments.

The near coastal component of EMAP consists of four ecosystem categories: estuaries, wetlands, coastal waters, and Great Lakes. A demonstration project will be conducted in the estuaries of the Virginian Biogeographic Province (Cape Henry, Virginia, to Cape Cod, Massachusetts) in 1990, followed by a full-scale national implementation.

Quality Assurance is an integral part of EMAP. The program will employ EPA's Data Quality Objective approach to ensure that the type, amount, and quality of data that are collected are adequate to meet program goals.

Introduction

Environmental regulatory programs have been estimated to cost more than $70 billion annually, yet the means to assess their effectiveness in controlling or mitigating environmental problems over the long term do not exist. While regulatory programs are based upon our best understanding of environmental processes at the time of their development, it is critical that long-term monitoring programs be in place to confirm the effectiveness of these programs in achieving the environmental goals and to corroborate the science upon which they are based.

---

* in collaboration with P.F. Hollarnd, Versar Inc., Columbia, Maryland 20145; R.M. Valente and K.J. Scott, Science Application International Corporation, Narragansett, Rhode Island 02882; S.C. Schimmel, U.S. Environmental Protection Agency, Narragansett, Rhode Island 02882; J.E. Pollard, Lockheed Engineering & Science Company, Las Vegas, Nevada 89119

Affirming the existence of a major gap in our environmental data and recognizing the broad base of support for better environmental monitoring, the EPA Science Advisory Board recommended in 1988 that EPA initiate a program that would both monitor ecological status and trends, as well as develop innovative methods for anticipating emerging problems before they reach crisis proportions. EPA was encouraged to become more active in ecological monitoring because its regulatory responsibilities require quantitative, scientific assessments of the complex effects of anthropogenic activities on ecosystems. The Environmental Monitoring and Assessment Program (EMAP) is EPA's response to all of these recommendations.

EPA's Office of Research and Development, in concert with several other federal agencies, is developing EMAP to determine the current status, extent, changes and trends in the condition of the nation's ecological resources. When fully implemented, EMAP will be able to respond to the following questions:

- What proportion of the nation's ecological resources are degrading or improving, where, and at what rate?
- What are the likely causes of the observed degraded conditions?
- What is the current status, extent, and geographic distribution of our ecological resources?
- Are control and mitigation programs effective in maintaining or improving the quality of the resources?

Objectives of EMAP

To provide the information necessary to address the above questions, EMAP has the three following objectives:

1. To estimate the current status, extent, changes, and trends in indicators of the nation's ecological resources on regional and national scales with known confidence limits;
2. To monitor indicators of pollutant exposure and habitat condition, and to seek correlative relationships between anthropogenic stresses and ecological condition that identify possible causes of adverse effects; and
3. To provide periodic statistical summaries and interpretive reports on ecological status and trends to the EPA Administrator and the public.

While the goal is to establish EMAP in all of the ecosystem types, the initial emphasis is on testing and implementing the program in the near coastal estuarine systems. A demonstration project for estuaries will be conducted in 1990 in the Virginian Biogeographic Province (Cape Henry to Cape Cod). This demonstration project will be used to design the full-scale national implementation of the program.

EMAP Near Coastal Approach

Assessing the status and trends in the condition of the nation's near coastal ecological resources requires data collected in a standardized manner, over large geographic scales, for long periods of time. Such assessments cannot be accomplished by aggregating data from the many individual, short-term monitoring programs that have been conducted in the past or are being conducted currently in the near coastal environment. Differences in the parameters measured, the collection methods used, timing of sample collection and individual program objectives severely limit the value of historical monitoring data for conducting regional assessments (Beanlands and Duinker, 1983, 1984; Wolfe et al., 1987). In addition, most historical and ongoing programs sample at specific locations, and the degree to which these data are representative of the overall system is rarely, if ever, known.

EMAP-Near Coastal proposes to overcome the problems discussed above by monitoring a defined set of parameters (i.e., indicators of near coastal environmental quality) using standardized sampling methods and an unbiased, probability-based sampling design on the spatial scale of biogeographical provinces, over a period of decades. These characteristics distinguish EMAP-Near Coastal from other monitoring programs and form the basis for preparing regional and national scale assessments that are needed to address environmental issues of the 1990s and beyond (Reilly, 1989; Thomas, 1988a, 1988b). In addition, these characteristics will allow EMAP-Near Coastal to provide: the information needed to evaluate and define the representativeness of existing monitoring data and sampling locations; an objective basis for incorporating the available historical data into the regional and national assessments; and a statistically reliable monitoring network and set of sampling protocols upon which local or system-specific monitoring programs can build.

Quality Assurance

EMAP-Near Coastal will use in excess of 30 individuals to collect samples and at least four different laboratories to process samples for the 1990 Virginian Province Demonstration Project. As the program is implemented in other provinces, the number of individuals and organizations involved in data collection and other aspects of the program will expand proportionately. Experience indicates that monitoring programs that involve a number of laboratories frequently encounter problems in obtaining data that are comparable among the many individuals and laboratories involved. Such problems usually result because, in the haste to initiate the collection program, an

evaluation of the comparability of the measurement methods and sampling designs to be used by each laboratory are not conducted.

Because the need for quality assurance (QA) programs increases as the complexity of the measurements and the number of organizations involved increases, a QA program will be implemented from the outset by EMAP to ensure that the data produced by the many groups are comparable. This program will also ensure that data are of known and acceptable quality. This program will consist of two distinct but related sets of activities: quality control and quality assessment.

Quality control includes design, planning, and management actions to ensure that the right types and amount of data are collected in the manner required to address study objectives. Examples of some quality control activities that will be employed by EMAP-Near Coastal are: development of standardized sample collection and processing protocols, and the requirement for specific levels of training before technicians are allowed to perform these activities. The goals of quality control procedures are to ensure that: the specified processing and analysis techniques are applied consistently and correctly; the number of lost, damaged, and uncollected samples are minimized; the integrity of the data record is maintained and documented from sample collection to entry into the data record; data are comparable with similar data collected elsewhere; and study results can be reproduced.

Quality assessment activities will be implemented to quantify the effectiveness of the quality control procedures. These activities ensure that the degree of measurement error and bias associated with the project are identified, accounted for, and eliminated (if practical). Quality assessment consists of both internal and external checks including repetitive measurements, internal test samples, interchange of technicians and equipment, use of independent methods to verify findings, exchange of samples among laboratories, use of standard reference materials, and audits (Taylor, 1987; EPA, 1985).

Data Quality Objectives

While QA is a necessary part of any sampling program, defining the proper level of QA to institute is difficult. If QA is defined too rigorously, it can quickly consume a disproportionate share of resources. If defined too leniently, the data collected may be of insufficient quality to meet program objectives. Within EMAP, the balance between cost and uncertainty will be established through the Data Quality Objectives (DQO) process (Fig. 1). The DQO process functions by specifying, early in the project, the questions to be addressed and the degree of uncertainty that decision makers who will use the data are willing to accept. These statements are translated into definitive, quantitative requirements for data accuracy, precision, representativeness, comparability, and completeness that are necessary to answer the specific questions identified at the

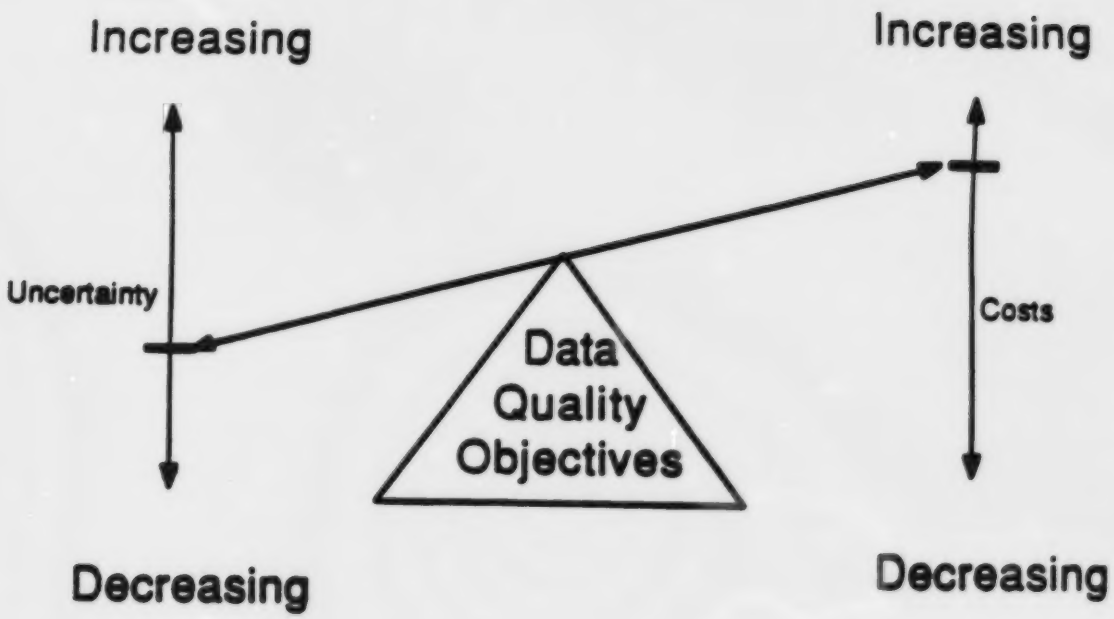


FIGURE 1. ROLE OF DATA QUALITY OBJECTIVES IN OBTAINING A BALANCE BETWEEN AVAILABLE RESOURCES AND THE LEVEL OF UNCERTAINTY.

onset. These requirements, as well as the procedures that will be implemented to meet them, are defined in a Quality Assurance Project Plan (QAPP) that must be approved by all levels of project management before the first sample is collected.

Developing DQOs is a multi-stage iterative process that involves individuals at all levels of the project. The first stage is initiated by the manager or decision maker, who identifies the central question to be addressed and the degree of acceptable uncertainty associated with the answer. In identifying acceptable uncertainty, the manager must weigh the cost of collecting samples against the "cost" of reaching incorrect decisions based on the sampling effort. The second stage is conducted by the project scientific staff, who formulate a sampling strategy for addressing the question, and then estimate the cost of developing an answer with the satisfactory level of precision. If the cost estimates are acceptable to the decision maker, then the project proceeds to the third stage, in which the technical staff develops quality control and quality assessment procedures for each aspect of the program (e.g., field collection, laboratory analysis and processing, data management, analysis) that are consistent with the desired level of quality. If cost estimates are too high, then the scientific staff and the decision makers jointly modify the design and expectations of the proposed program until a proper balance of costs and uncertainty is achieved.

Two sources of error are considered in establishing DQOs: sampling error and measurement error. Sampling error is the difference between the sampled value and the true value and is a function of natural spatial and temporal variability and sampling design. The temporal variability relevant to EMAP is that which occurs within the sampling period. Measurement error is the difference between true sample values and the reported value, and can occur during the act of sampling, sample treatment, sample splitting, sample analysis, data entry, data base manipulations, etc. While "good" data are available to estimate measurement error for all of the parameters that will be used in EMAP-Near Coastal, data for estimating sampling error are either unavailable or unaccessible for many, if not most, of the variables to be measured. Acceptable estimates of variability are unavailable because EMAP is the first program to measure most of these parameters on a regional scale using standardized methods and a probability-based sampling design.

Good estimates of temporal and spatial variability are essential to the DQO process since they are required for quantifying the degree of uncertainty that will be produced by the sampling design. Without them, the feedback loop in which the scientific staff provides the decision makers with an estimate of cost for a desired level of uncertainty cannot take place. For this reason, DQOs will not be implemented in the 1990 Demonstration Project. Rather, one of the goals of the demonstration project will be to gather the data to establish DQOs for the implementation phase in subsequent years. Instead, the demonstration project will be implemented using Measurement Quality Objectives (MQOs). MQOs establish acceptable levels of uncertainty for each measurement process, but differ from DQOs in that they are not combined with sampling error to estimate programmatic uncertainty. MQOs have been

established by obtaining estimates of achievable data quality based on manufacturer specifications, the judgement of knowledgeable experts, and available literature information. Each measured parameter will have an associated MQO established for five attributes of data quality: representativeness, completeness, comparability, accuracy, and precision (Stanley and Verner, 1985). In subsequent years, DQOs will be developed to replace the MQOs. Data quality attributes are defined below along with the MQO established for each measured parameter within EMAP-Near Coastal.

- Representativeness is the degree to which the data represent a characteristic of a population parameter. In EMAP-Near Coastal, representativeness is most germane to the proper siting of sampling stations, and the MQO will be to ensure that all samples with the exception of fish trawling are within 100 meters of the planned sampling site. Fish trawling should occur within 500 meters of the planned site.

- Completeness is a measure of the amount of valid data (i.e., not associated with some criteria of potential unacceptability) collected from a measurement process compared to the amount that was expected to be obtained. The MQO completeness criteria for EMAP-Near Coastal will range from 75 to 90 percent depending on the measurement process.

- Comparability is defined as "the confidence with which one data set can be compared to another" (Stanley and Verner, 1985). Comparability of reporting units and calculations, data base management processes, and interpretative procedures must be assured if the overall goals of EMAP are to be realized. The MQO for the 1990 Virginian Province Demonstration Project is to use accepted methods and to generate a high level of documentation to ensure that future EMAP efforts can be made comparable.

- Accuracy is defined as the difference between a measured value and the true or expected value, and represents an estimate of systematic error or net bias (Kirchner, 1983).

- Precision is defined as the degree of mutual agreement among individual measurements, and represents an estimate of random error (Hunt and Wilson, 1986).

Quality Assurance Reports

The first Annual Statistical Summary for EMAP-Near Coastal is scheduled for June 1991 after completion of the 1990 Virginian Province Demonstration Project. This report will summarize the measurement error estimates for the various data types. Precision, accuracy, comparability, completeness, and representativeness of the data collected during the demonstration project will be summarized in this document and

detection limits reported. After implementation of the near coastal component of EMAP, progress will be reported on an annual basis. Interpretive Assessment Reports will be prepared every four years. In addition, Special Scientific Reports will be produced periodically to address concerns raised about the program such as the ability of the design to detect trends or to properly categorize subnominal values. The data quality attributes of precision, accuracy, comparability, completeness, and representativeness will also be provided for each of the reports.

Acknowledgements

Although the work described in this article was supported by the U.S. Environmental Protection Agency, it has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency; no official endorsement should be inferred.

References

Beanlands, G.E. and P.N. Duinker. 1983. An ecological framework for environmental impact assessment in Canada. Institute for Resource and Environmental Studies, Dalhousie University. Halifax, Nova Scotia.

Beanlands, G.E. and P.N. Duinker. 1984. Lessons from a decade of offshore environmenal impact assessment. Ocean Management 9:157-175.

Environmental Protection Agency. 1985. Standard Operating Procedures for Conducting Surplus and Sample Blank Audits. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada, EPA/600/4-85/003. 71pp.

Hunt, D.T.E., and A.L. Wilson. 1986. The Chemical Analysis of Water: General Principles and Techniques. 2nd ed. Royal Society of Chemistry, London, England. 683 pp.

Kirchner, C.J. 1983. Quality Control in Water Analysis. Environ. Sci. and Technol. 17(4):174A-181A.

National Research Council. 1990. Managing Troubled Waters: The Role of Marine Environmental Monitoring. National Academy Press, Washington DC.

Reilly W.K. 1989. Measuring for Environmental Results. E.P.A Journal 15(3):2-4.

Stanley, T.W., and S.S. Verner. 1983. Interim Guidelines and Specifications for Preparing Quality Assurance Project Plans. EPA/600/4-83/004. U.S.

Environmental Protection Agency, Washington, D.C.

Taylor, J.K. 1987. Quality Assurance of Chemical Measurements. Lewis Publishers, Inc., Chelsea, Michigan. 328pp.

Thomas, L.M. 1988a. The E in Environmental Protection Agency. E.P.A. Journal 14(1):3-4.

Thomas, L.M. 1988b. Environmental Regulation Challenges We Face. E.P.A. Journal 14(2):2-3.

Wolfe, D.A., M.A. Champ, D.A. Flemer, and A.J. Mearns. 1987. Long-term biological data sets: their role in research, monitoring, and management of estuarine and coastal marine systems. Estuaries 10:181-193.

IN-HOUSE REFERENCE MATERIALS AS A MEANS TO QA: THE CWS EXPERIENCE

Richard Turle
National Wildlife Research Centre
Canadian Wildlife Service
Environment Canada
Ottawa, Ontario, K1A 0H3

Abstract

For many years the Canadian Wildlife Service has used in-house reference materials for monitoring QA performance, both by contractors and in-house, primarily for organochlorine/PCB determinations. The reasons for this approach are given. Analysis of the data received from check sample programs and other exchanges have allowed usable values to be assigned. Analysis of in-house QC charts have been used to determine causes of variability and to predict the limitations of the analytical method. The application of this information to clients is discussed.

Introduction

The main purpose of long-term monitoring of toxic chemical residues is to establish changes in concentration with respect to time. The analytical chemist may not be concerned with identifying the physical sources of the contaminants or to interpret the trends in terms of biological effects, populations, etc., but the chemist is charged with producing analytical results that are accurate, precise and comparable to previous results, that is, results without bias with respect to the true value, however obtained. The reasons for this are obvious; compounds may be mis-identified and trends in the levels of contaminants may be mis-interpreted. Also, it may not be possible to see the effects of regulatory action on chemicals entering the environment, if imprecision masks changes in actual levels.

Because of this need to produce accurate data, the chemist is very concerned with Quality Assurance (QA) and all that it entails. Two successful aspects of the QA program of the Canadian Wildlife Service (CWS) long-term monitoring studies have been the ability to retrieve samples from a specimen bank and the use of reference materials. The retrieval of samples ensures that the analyst can determine levels of compounds that he was unable to determine in the past and, perhaps as importantly, ensures that changes in methods are not obscuring long-term trends. The

establishment of reference materials enables a direct comparison between methods, analysts and laboratories over a long time frame.

Since 1969, CWS has been involved in measuring organochlorine residues in wildlife. During this period one basic method has been used, but one which has been continually improved. This method is based on the fact that most samples are tissues with a moderate lipid content. Simply, the sample is ground with anhydrous sodium sulphate, until dry, placed in a glass chromatographic column and the lipid extracted with 50% methylene chloride in hexane. A portion is removed for lipid determination and the balance separated on a Florosil column into 3 fractions, using 0, 15 and 50% methylene chloride in hexane. The extracts are analyzed by GC-ECD on a 60M x 0.25 mm fused silica column.

CWS Reference Materials

The CWS approach to both in-house and contract analysis has been based on Reference Materials (RMs). Why? With wildlife samples it is not possible to prepare control samples or "spike" into a sample because often the compounds are not available or, in the case of PCBs, the pattern is different from that found in Aroclor mixtures due to the metabolic process. Standard Reference Materials, such as those produced by NIST would be better but only one, SRM 1588 Cod Liver Oil is suitable. Because it is fish oil it does behave in our method like animal lipids. A further disadvantage of this SRM, besides cost, is that the published values do not include very many PCB congeners.

The first SRM was produced in 1979, using Herring Gull eggs from Lake Erie. A second was produced in 1980 with eggs collected from Lake Ontario, primarily because the PCB levels in the 1979 RM were too high for routine analysis. A third RM was produced in 1986 for use in dioxin analysis. We have collected more eggs this last year for developing another egg RM so we can extend the life of the 1980 RM. Because CWS has the responsibility for monitoring contaminant levels in wild foods, such as hunted migratory game birds, we will develop in 1990 a Breast Muscle RM, using ducks caught in nets in Lake Ontario. Although there are good NIST SRMs for metals, we will use these same birds for development of liver and kidney RMs for metals. As ecological monitoring changes from measurement of persistent chemical contaminants to measuring biomarkers (i.e. biochemical indicators of exposure) we will attempt to develop suitable in-house RMs.

Obviously one question is how good are our RMs. In order to find out, we have freely exchanged our RMs with other laboratories (Turle et al., 1988). Table 1 gives the results for the 1980 RM. For this RM we have a considerable number of results from ORF which was the main CWS contractor for a number of years. Comparisons for PCB congeners have been more difficult to make than for organochlorine residues. Currently, we have a contract with the Great Lakes Institute of the University of

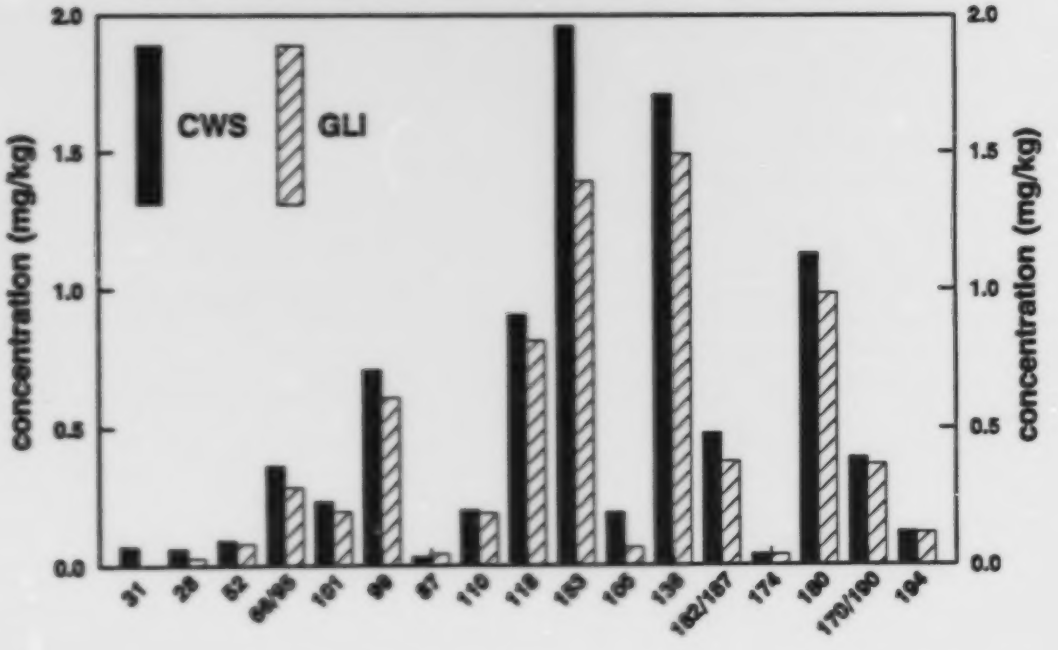
Table 1: Inter-laboratory comparison of CWS 1980 Reference Material

Residue Determinations	CWS	CV	ORF (6)*	CV	Chemex (8)	CV	Manitoba (5)	CV	GLI (9)	CV
HCB	0.073	6	0.073	6	0.060	22			0.066	45
pp'DDE	5.00	5	4.59	5	6.35	12	3.99	4	4.36	
Aroclor 1:1	23.4	7	23.4	7	22.4	11	22.4	3	22.1	4
Mirex	0.144	5	0.144	5	0.202	20	0.180	4	0.131	43
Oxychlordane	0.239	10	0.239	10					0.187	12
c-Chlordane	0.019	19	0.019	19	0.045	52			0.009	77
t-Nonachlor	0.023	49	0.023	49	0.108	54				
pp'DDD	0.010		0.010		0.013	67			0.013	70
pp'DDT	0.060	13	0.060	13	0.080	48	0.028	16	0.035	95
Hexachlor Epoxide	0.140	11	0.140	11	0.111	45	0.046	19	0.106	30
Dieldrin	0.276	13	0.276	13	0.295	14	0.300	2	0.215	44

All results expressed in mg/kg wet weight. Coefficient of Variation is expressed as a %. ORF, Ontario Research Foundation; Chemex, Chemex Labs Alberta (1984) Ltd; Manitoba, Manitoba Technical Services Branch laboratory; GLI, Great Lakes Institute, University of Windsor. Blanks indicate not analyzed.

* Study number in parentheses.

Fig. 1 1980 RM PCB Inter-Laboratory Comparison



Windsor for OC/PCB analysis. The comparison between CWS and GLI for the 17 PCB congeners they analyze is given in Figure 1. Generally, they are biased low but they did meet our overall criteria for acceptability of results for a contract laboratory (Turle et al., 1990). We believe that the inter-laboratory comparisons indicate that the values generated provide "usable" Reference Values for determining the performance of our and external laboratories. We would like to be able to validate our new RMs so we intend to be sending out samples of the new Herring Gull egg and Breast Muscle RM in the fall. Your participation would be welcome.

Of course, these results would be meaningless unless we could demonstrate that we could analyze SRMs well. The only SRM we have analyzed is the newly released Cod Liver Oil. Table 2 provides our results on this material as well as giving the NIST values. To date, we have not obtained information from other labs who have determined the specific PCB congeners in this SRM. If it does turn out that our results for certain congeners are incorrect, this may well lead us to adjust the corresponding values in the CWS RMs.

How are the RMs used on a daily basis? Each batch of samples contains 1 in 10 RM samples. These results are extracted from the analytical reports and compared with the long term means. A result for an individual parameter is accepted if it is within ± 2 s.d. of the long-term mean. If it is outside this criterion, the result is accepted if it is within ± 3 s.d. AND there has not been a non-complying result accepted for the previous 19 analyses. In other words, we accept that if the results on the RM are normally distributed 1 in 20 will exceed the ± 2. s.d criterion, but that only 1 in 95 should exceed ±3 s.d. However, if several parameters fall beyond the ±2 s.d. limit then the batch should be re-analyzed. Reasons for re-analysis are normally either the extraction for the sample was poor and all the parameters are too low or the calibration for the gas chromatograph is out for some reason. We now monitor the lipid levels in the RM to ensure consistency in extraction. This cannot be overemphasized as most of the comparisons and long-term trends for wildlife are made on a lipid weight basis.

Sources of Variation

For each parameter analyzed, mean, s.d., CV and control charts have been produced. Examination of the control charts indicated that many at high concentration showed very similar patterns (e.g. PCB 138 and PCB 153, Fig. 2). Other control charts for congeners at low concentrations (e.g. PCB 97) showed a very different pattern for each congener. If all the results for each congener are correlated against the results of the congener present in highest concentration, PCB 153, there is a decline in the correlation coefficient with concentration (Fig. 3). Generally this can be explained if we presume that two processes are operating. At higher concentrations variations in the process which affect the sample as a whole predominate. These variations are the

Table 2: **CWS results on NIST SRM 1588 Organics in Cod Liver Oil**

Residue		NIST Ref	NIST 2 s.d.	NIST CV	CWS mean	CWS 2 s.d.	CWS CV
1245-TCB					0.002	0.010	0.52
1234-TCB					0.014	0.008	0.41
QCB					0.010	0.005	0.23
HCB		0.148	0.021	.05	0.137	0.015	0.76
OCS					0.017	0.001	0.07
t-NONACHLOR		0.209	0.011	0.55	0.256	0.033	1.64
pp'DDE		0.641	0.062	3.10	0.596	0.035	1.75
pp'DDT		0.529	0.045	2.25	0.587	0.104	5.19
PHOTOMIREX					0.007	0.003	0.14
MIREX					0.009	0.002	0.08
alpha-BHC		0.086	0.019	0.95	0.055	0.002	0.09
beta-BHC					0.004	0.000	0.01
g-HCH					0.015	0.003	0.14
OXYCHLORDANE					0.033	0.006	0.32
t-CHLORDANE		0.05	0.013	0.65	0.038	0.011	0.56
c-CHLORDANE		0.158	0.008	0.40	0.120	0.015	0.76
pp'DDD		0.277	0.015	0.75	0.225	0.059	2.94
c-NONACHLOR					0.070	0.011	0.54
CB	31				0.003	0.010	0.50
	28				0.018	0.010	0.52
	52				0.071	0.004	0.22
	49				0.027	0.006	0.28
	47				0.184	0.439	21.93
	44				0.025	0.004	0.21
	42				0.008	0.020	1.00
	64				0.004	0.000	0.02
	70				0.015	0.006	0.31
	66				0.086	0.008	0.40
	60				0.040	0.008	0.39
	101	0.129	0.005	0.25	0.147	0.013	0.67
	99				0.231	0.020	1.00
	97				0.026	0.009	0.45
	87				0.028	0.056	2.79
	110				0.075	0.018	0.91
	118				0.218	0.047	2.33
	146				0.064	0.008	0.38
	153	0.276	0.04	2.00	0.305	0.041	2.05
	105				0.059	0.006	0.31
	141				0.031	0.007	0.36
	137				0.028	0.013	0.64
	138	0.261	0.029	1.45	0.292	0.056	2.79
	158				0.012	0.003	0.14
	129				0.023	0.004	0.21
	182				0.061	0.017	0.87
	183				0.035	0.010	0.50
	128				0.292	0.104	5.21
	185				0.001	0.005	0.25
	174				0.023	0.005	0.27
	171				0.012	0.024	1.22
	200				0.046	0.083	4.13
	172				0.003	0.010	0.48
	180	0.107	.004	0.20	0.128	0.049	2.45
	170	0.045	0.005	0.25	0.062	0.021	1.05
	201				0.020	0.006	0.30
	203				0.014	0.004	0.22
	195				0.003	0.008	0.38
	194				0.016	0.010	0.48
	206				0.007	0.016	0.78
	Σ PCB				2.750	0.328	16.42

CWS mean values based on 4 determinations. Fraction 3 (heptachlor exporide and dieldrin) was not analyzed due to chromatographic problems. (Turle et al., 1990)

Fig. 2 Examples of Quality Control Charts

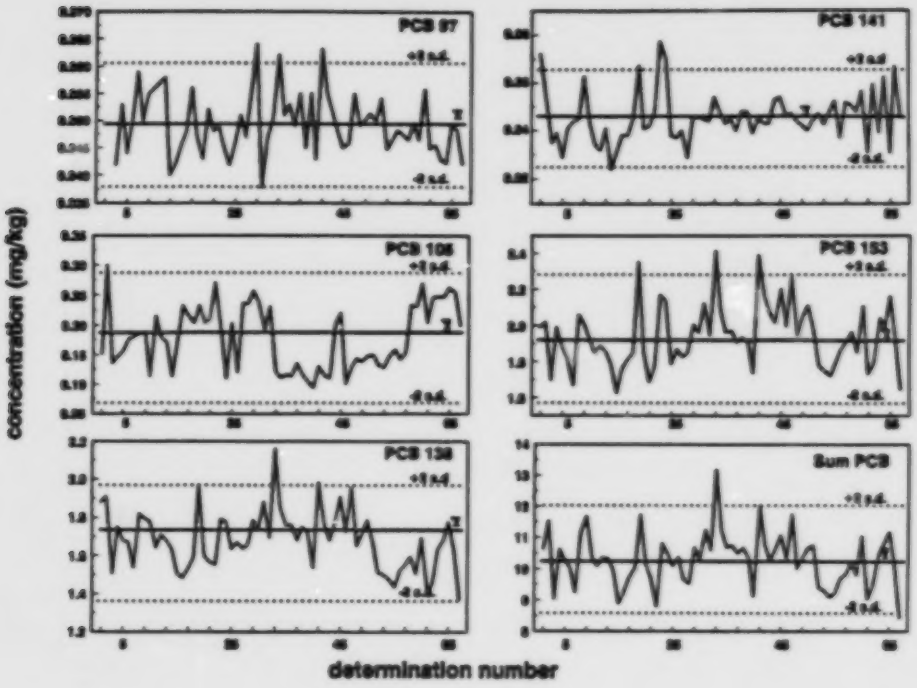
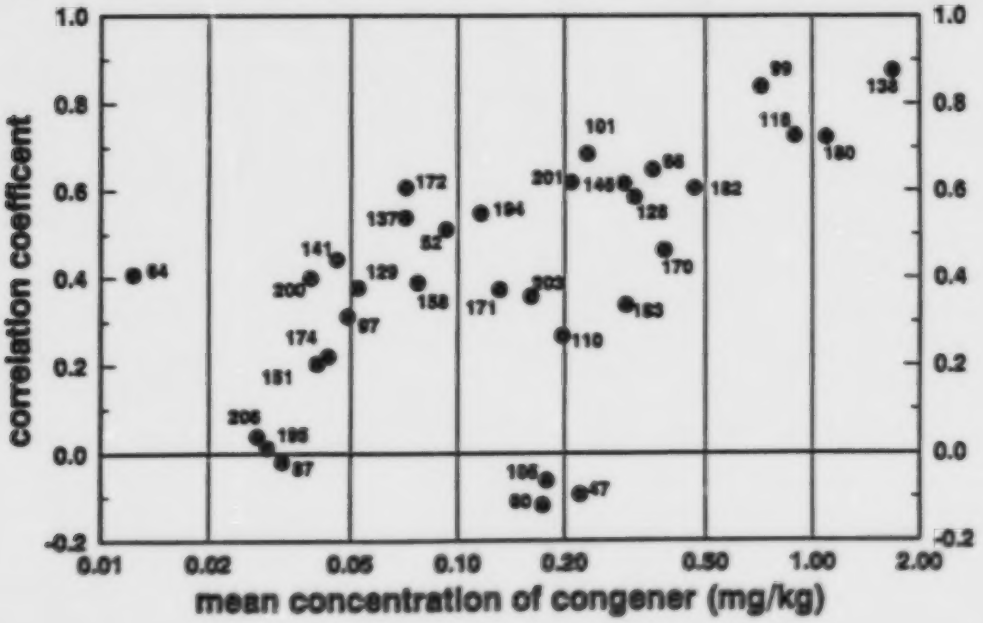


Fig. 3 Correlation of PCB congener results with PCB 153 results



steps in the procedure such as weighing, extraction, evaporation and dilution. In the determination step, variations in the calibration or in the volume actually injected are those most likely to affect the sample as a whole. At very low concentrations those factors causing variation which affect each congener differently predominate. For example, at the determination stage the concentration and type of interference is more likely to add to or reduce the real concentration. Further, the ability of the integrator to determine the exact start or end of a peak declines with concentration. Often because of poor resolution, part of the area integrated for one congener will be included in an adjacent peak. Instrument noise is more likely to affect chromatographic processing of congeners near the instrument detection limit than congeners present at high concentration. We postulate that the similarity in the control chart patterns for PCBs such as 138 and 153 compared to the difference in patterns between congeners in low concentration such as PCB 97 and 141 can thus be explained.

An examination of the plot indicates that there are several congeners which are not correlated. These are 47, 60 and 105. The peak that has been identified as PCB 47 is often subject to interference. Indeed with samples of Σ PCB concentrations of less than 1 mg/kg this peak constitutes sometimes 30% of the Σ PCB result. PCB 60 co-elutes with PCB 56 and the peak shape is not always consistent. PCB 105 is never very well resolved from 153 despite its relatively high concentration. The resolution between this pair of PCBs is a good indicator of column efficiency. PCB congeners 87, 195 and 206 are the least correlated because these are the congeners for which we have the most incomplete data sets. In contrast, PCB 64 is relatively well correlated despite its very low concentration in the RM. This is hard to explain since it co-elutes with PCBs 41 and 71. It is possible that since this peak is well resolved, it is less subject to the variations which affect other peaks.

Total PCB concentrations can be obtained by summing all the concentrations of the individual congeners. From the above arguments it is no surprise that the CV for Σ PCB (8.2 %) is slightly lower than that obtained by the previous one-peak method based on PCB 138 (8.9 %). This is despite the Σ PCB value being obtained by summing 34 results. Since PCBs 99, 118, 138, 153 and 180 constitute 61 % of the Σ PCB value and the 16 PCBs with concentrations present at greater than 0.2 mg/kg constitute 88 % of the Σ PCB value, we would anticipate those factors which affect the sample as a whole to predominate in determining the CV. Some of the variation in concentration in one congener caused by peak processing is effectively cancelled out by another, thereby reducing the CV lower than that obtained for the congeners in highest concentration. Thus the CV of the Σ PCB reflects the variation inherent in the extraction, separation and measurement steps of the method as a whole. It is not likely that the CV for a manual method such as this could be reduced except by automation. This fact, that the CVs are approximately the same, is of concern when calculating conversion factors to calculate Σ PCB values based on existing Aroclor 1:1 1245/1260 data.

Changes of CV with Concentration

For some time in assessing contracts, CWS has used the concept of CV_{max} to allow for the systematic increase of CV as concentration declines (Turle et al., 1987, 1988). The CV_{max} criterion is based on the work of Frehse and Timme (1980). Briefly, it has been shown that, as analyte concentration declines the CV increases. From this they developed a concept of "first category measured value curves". This can be extended to the general use equation:

$$CV_{max} = CV_o \ (x_o/x)^{\log f} \ (\%)$$

where CV_{max} = maximum permissible CV
CV_o = CV at the detection limit
x = mean of replicate analysis of blind sample
x_o = detection limit
f = the factor by which CV_{max} diminishes per order of magnitude.

For example, if $f = 2$ and $CV_o = 100$, the equation reduces to:

$$CV_{max} = (x_o/x)^{0.3} \cdot 100 \ (\%)$$

For pesticide residue analysis, the form of the equation is:

$$CV_{max} = 13.7 \ x^{-0.431} \ (\%)$$

based on $x_o = 0.01$ mg/kg, $CV_o = 100\%$ and $CV_{max} = 15\%$ at 1mg/kg. This is equivalent to $f = 2.7$. As an example, for a mean of 0.0895 mg/kg, CV_{max} is 38.8%. These equations allow CV_{max} to increase as concentrations decline in a systematic manner. A practical lower limit of CV_{max} is 15%.

Since we had access to the data on PCB congeners, we attempted to verify if the equation we had used was valid. For PCBs the CV at any given concentration can be expressed as

$$CV = 11.33 \ x^{-02542}$$

Tables 3 and 4 illustrate actual CV's for PCBs and other organochlorine residues, respectively, in the 1980 Herring Gull RM.

Table 3: Statistics for PCB congeners obtained for the 1980 Herring Gull RM

Congener	Mean mg/kg	s.d. mg/kg	CV% %	n	Lowest mg/kg	Highest mg/kg	IDL[1] mg/kg
47	0.22136	0.08855	40.0	17	0.0657	0.4546	0.002
52	0.09266	0.02039	22.0	56	0.0480	0.2010	0.007
60	0.16882	0.07363	43.6	35	0.0000	0.3480	0.009
64	0.01228	0.00182	14.8	46	0.0090	0.0160	0.001
66	0.35519	0.03936	11.1	65	0.2325	0.4630	0.005
87	0.03096	0.01708	55.2	29	0.0000	0.0670	0.004
97	0.04903	0.00557	11.4	61	0.0380	0.0640	0.005
99	0.71624	0.06580	9.2	67	0.5797	0.9040	0.006
101	0.23171	0.02060	8.9	63	0.1880	0.3040	0.005
105	0.17832	0.05295	29.7	66	0.0950	0.2700	0.003
110	0.19769	0.03187	16.1	67	0.0190	0.2750	0.004
118	0.89086	0.08965	10.1	67	0.6286	1.1200	0.004
128	0.31530	0.05418	17.2	67	0.1603	0.4240	0.004
129	0.05225	0.00758	14.5	67	0.0350	0.0698	0.002
137	0.07107	0.01059	14.9	67	0.0390	0.1040	0.003
138	1.68193	0.15050	8.9	67	1.3740	2.1600	0.004
141	0.04971	0.03399	68.4	66	0.0240	0.3087	0.003
146	0.29419	0.03369	11.5	67	0.2140	0.3880	0.003
151	0.03620	0.01377	38.0	10	0.0000	0.0470	0.004
153	1.94189	0.17253	8.9	67	1.6300	2.4100	0.003
158	0.07712	0.01003	13.0	63	0.0521	0.0986	0.001
170	0.38281	0.07154	18.7	67	0.2813	0.7360	0.003
171	0.13098	0.03659	27.9	62	0.0078	0.2362	0.002
172	0.07132	0.01595	22.4	67	0.0040	0.1080	0.002
174	0.04245	0.01083	25.5	59	0.0000	0.0890	0.004
180	1.09410	0.12421	11.4	67	0.8714	1.4600	0.003
182	0.46544	0.05885	12.6	67	0.2766	0.6350	0.003
183	0.29849	0.06579	22.0	67	0.1210	0.5626	0.003
194	0.11601	0.02231	19.2	67	0.0780	0.1770	0.003
195	0.02910	0.01372	47.1	25	0.0090	0.0550	0.002
200	0.03848	0.00848	22.0	49	0.0137	0.0600	0.003
201	0.20905	0.02935	14.0	67	0.1251	0.2950	0.004
203	0.16077	0.03425	21.3	67	0.1143	0.3790	0.002
206	0.02734	0.01334	48.8	19	0.0134	0.0780	0.003

[1] Instrument Detection Limit based on an area reject of 5555 μV/s for an HP 3392 integrator.

Table 4: Statistics for Organochlorine Residues in 1980 Herring Gull RM

Residue	Mean mg/kg	s.d. mg/kg	CV %	n	Lowest mg/kg	Highest mg/kg	IDL[1] mg/kg
QCB	0.01850	0.02390	129.2	15	0.0077	0.0969	0.0030
HCB	0.08189	0.00906	11.1	96	0.0606	0.1100	0.0020
OCS	0.02789	0.00294	10.6	36	0.0201	0.0337	0.0020
t-NONACHLOR	0.12238	0.03313	27.1	96	0.0875	0.2580	0.0030
pp'DDE	4.99575	0.44700	8.9	96	4.0340	6.2000	0.0040
pp'DDT	0.08889	0.01603	18.0	96	0.0522	0.1260	0.0050
PHOTOMIREX	0.06212	0.00745	12.0	86	0.0488	0.0945	0.0050
MIREX	0.15687	0.02316	14.8	96	0.1340	0.2686	0.0030
alpha-BHC	0.08972	0.06835	76.2	86	0.0018	0.2686	0.0005
beta-BHC	0.02249	0.04115	182.9	96	0.0069	0.1900	0.0002
g-HCH	0.00955	0.00168	17.6	65	0.0052	0.0140	0.0005
OXYCHLORDANE	0.21155	0.02260	10.7	96	0.1593	0.2630	0.0003
t-CHLORDANE	0.20989	0.02407	11.5	64	0.0000	0.2630	0.0003
c-CHLORDANE	0.02141	0.02031	94.9	86	0.0111	0.2020	0.0003
pp'DDD	0.01409	0.01997	141.7	96	0.0063	0.2020	0.0007
c-NONACHLOR	0.05774	0.00708	12.3	89	0.0423	0.0770	0.0003
H.E.	0.13396	0.01370	10.2	96	0.1060	0.1690	0.0010
DIELDRIN	0.26885	0.02986	11.1	96	0.2362	0.3640	0.0010
AR 1:1[2]	23.275	2.226	9.6	96	20.2600	29.60	0.0547
AR 1260	9.676	1.234	12.8	89	8.5340	12.78	0.0258
%FAT	8.427	0.569	6.7	38	7.87	9.58	7.15
%WATER	77.195	0.614	0.8	38	74.92	78.44	76.19

[1] Instrument Detection Limit based on an area reject of 5555 μV/s for an HP 3392 integrator.
[2] Aroclor 1254/1260 1:1 mixture.

Figure 4 indicates that our criterion for CV_{max} encompasses all but two congeners. Other work has indicated that the peak assigned to PCB 47 is probably an interfering compound and the peak assigned to PCB 60 coelutes with PCB 56 and is not consistent in shape. The best fit of the data was calculated and was found to be similar in form but with slightly different coefficients. A similar exercise was performed on the organochlorine residues. The problem with this data set is that there is not such a wide range of concentrations as found with the PCB congeners. However, the best fit again produced a similar curve with coefficients very similar to that found for PCBs.

$$CV = 11.56\ x^{-0.303}$$

Thus for this method, it is possible to produce an overall equation linking CV and concentration.

$$CV = 10.286\ x^{-0.316}$$

It would be interesting to perform a similar exercise with a different methodology used for a wide range of analytes.

These relationships can be used by the recipients of the data to decide if the variation among results is likely to be due to the analytical variation or due to the inherent variation in the samples. This is not a trivial consideration as often the decision as to whether there are significantly different levels of a toxic chemical between two populations are based on real, often imprecise data. If we assume that our method has a precision of 8% at 1 mg/kg (which is probably the lowest obtainable limit) then at 1 μg/kg the CV will be about 80%. If our method is less precise, at 1 mg/kg then the ability to provide usable data declines more rapidly with decreasing concentration. Many pesticide methods have a CV of around 25% at 1 mg/kg. At 1 μg/kg the CV becomes about 500%. If the acceptable range of possible values for a sample is considered ± 2 s.d. of the true value, then if this is expressed in terms of concentration for a method having a CV of 25% at 1mg/kg the range of results becomes 1 mg/kg. Even for a CV of 8%, ± 2 s.d constitutes a range of 0.3 mg/kg at 1 mg/kg.

Conclusion

Finally, in concluding, it is possible to produce accurate and precise results for a complicated analytical method as described. However, it is not achieved without considerable effort. CWS has found that in-house RMs are a very useful method for maintaining quality and for assessing the performance of external laboratories. Thus the following quote from John Ruskin seems appropriate. "Quality is never an accident; it is always the result of intelligent effort."

Acknowledgements

Most of the results presented here represent the fine analytical chemical skills of Henry Won. I also wish to thank Ross Norstrom for his many ideas and Brian Collins for his help with statistics. Finally, I must thank those analysts in various laboratories who have provided the results for the inter-laboratory comparisons.

References

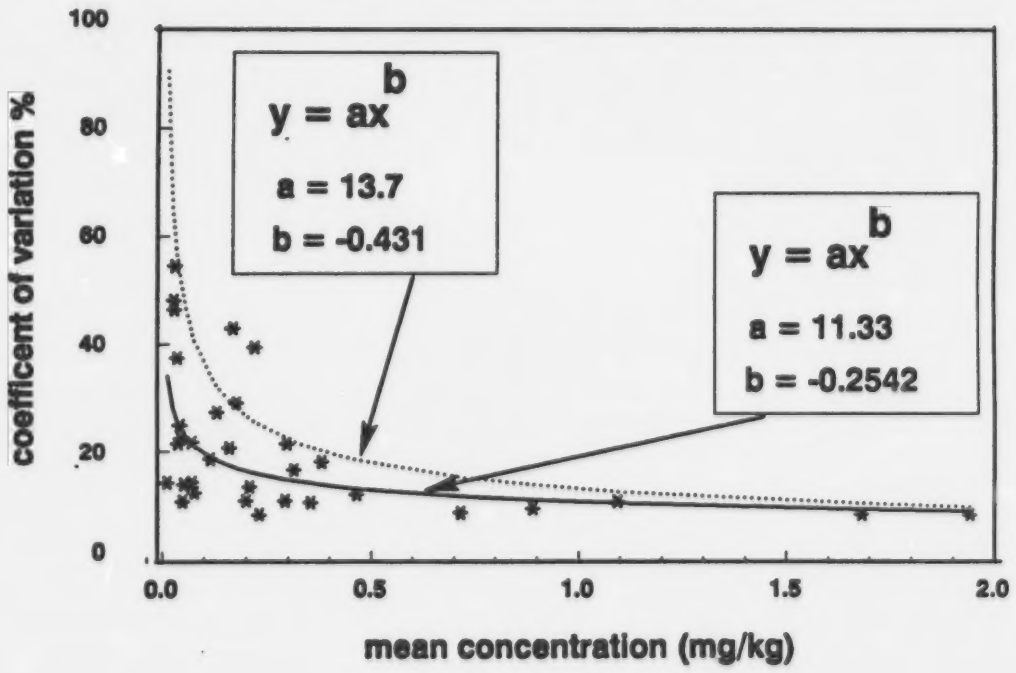
R. Turle, R.J. Norstrom and H.T. Won. 1988. Quality Assurance as Applied to Long-term Monitoring of Chemical Residues and Specimen Banking, In Progress in Environmental Specimen Banking, Natl. Bur. Stand. (U.S.), Spec. Publ. 740, pp 171-183 (Apr. 1988)

R. Turle, R. J. Norstrom, H. Won and B. Collins. 1990. Precision in the determination of PCBs (paper in preparation).

R. Turle and R.J. Norstrom. 1987 The CWS Guideline to Practical Quality Assurance for Contracted Chemical Analysis, Technical Report Series No.21, Canadian Wildlife Service Headquarters.

H. Frehse and G. Timme. 1980. Residue Rev. 73: 27-47.

Fig. 4 CV versus Concentration for PCBs



TRACE ANALYSIS OF ATMOSPHERIC PRECIPITATION SAMPLES

W. M. J. Strachan,
Lakes Research Branch
Environment Canada
Burlington, Ontario, L7R 4A6

Abstract

Most of the persistent organic contaminants found in environmental samples from the Great Lakes reached there via long range transport through the atmosphere. Historically, analyses have focussed on the more readily determined levels in biota and sediments but present and future demands for modelling and prediction require that the basic compartments, particularly the mobile ones -- air and water -- be assessed.

Rain samples have been collected with equipment designed at CCIW which utilizes XAD-2 resin sorption to concentrate the sample analytes in the field. Discussion will focus on the concentration levels and the variability to be expected in such samples. Substances of particular concern include total PCBs (the future will require congener specific analyses), the hexachlorcyclohexanes (a-HCH and lindane), DDT residues, dieldrin, HCB and others. Levels in rain observed across Canada vary from <0.1 ng/L to nearly 30 ng/L and account for direct inputs to some of the lakes of literally tonnes/annum.

Introduction

In the Great Lakes, and elsewhere, the major mechanism by which many toxic chemical pollutants are introduced to the terrestrial and aquatic parts of the ecosystem is via atmospheric transport from remote distances (International Joint Commission, 1987). This applies to such notable contaminants as the polychlorinated biphenyls, DDT residues, dieldrin, lead and other substances. They reach the system from intercontinental distances, in some cases, and are deposited in the form of rain (and snow), dryfall particulate matter and vapour state exchange with water and soil/leaf material.

Quality assurance programmes, understandably for the Great Lakes, have focussed on relatively homogeneous samples such as water. To a lesser extent, biota and sediments have been the media but in most cases, it seems to this researcher that metals and inorganic anions have been the analytes. This may be the consequence of historic analytical detection limits which precluded working with the low levels of the persistent organic pollutants found in the basic media -- air, water and sediment/soil, and the suspended matter found in the first two. The present state of analytical art,

however, is much advanced over that of ten years ago and it is appropriate that analytical and quality assurance investigators turn their attention there. These sample media, and the properties of the individual chemicals, govern the distribution of chemicals in the environment and are therefore those which will determine how long the problems of such toxicants will last. It is appropriate, therefore that these sample matrices receive more attention than they have in the past. The subject of this paper relates to samples from atmospheric sources (and their attendant quality assurance aspects) but the other basic media are equally in need of investigation.

Atmospheric Sample Types

Precipitation, in this presentation, is taken very broadly to include all sample types which result in the transfer between the air and the terrestrial or aquatic compartments of the environment. In a practical sense, work has largely focussed on the air-water interactions and all comments are restricted to these two compartments unless noted. This limits the sample types to rain, dryfall air particles and vapour state. There is less information on the latter two; consequently, most comments are derived from experience with rain.

The samples for rain are usually large relative to traditional water samples -- tens of litres -- although this often corresponds to a few centimetres of wetfall which can be acheived in a single, heavy rainfall (Strachan and Huneault, 1984). Routine sample collection, however, usually composites a number of precipitation events collected unfiltered either on an XAD resin column or in a solvent. Samples may sit adsorbed to the column material for up to a month, are shipped via post, and processed in the laboratory at anything up to a month or two following receipt; they are refrigerated during this latter period although not for other times.

Samples for the air compartment are usually collected on large glass filters (for particulates) which are placed before a vapour state trap (polyurethane foam plug, XAD or Tenax trap, etc.). These two collectors are part of a "high volume" (Hi-Vol) sampler pumping at the rate of cubic metres per minute. There is a lot of uncertainty about whether pollutants sorbed on the particulates (trapped on the filter) are retained there or whether they desorb and increase the apparent atmospheric vapour state load (e.g., Konig et al., 1980; Broddin et al., 1980; Grosjean, 1983). The Atmospheric Environment Service (Canada) together with the Ontario Research Foundation (now ORTECH International) has developed a sampler which separates vapour and particle using the denuder principle (Lane et al., 1988). This involves the vapour being selectively removed by absorption into a Tenax/silicon gum coating on concentric glass tubes; the particles are trapped on a conventional filter backed up by a Tenax trap. The filter and backup are determined following refrigeration in the field and processing and workup in the laboratory usually within a few weeks of returning to the laboratory. Vapour state is found as the difference between total (no denuders) and the particulates as described.

In all cases, the analytes must be removed from the sample trap medium. Except for systems using solvents to concentrate in the field, this involves eluting with a solvent, subsequent cleanup and possible fractionation, and analysis by gas-chromatography. Because environmental samples are inherently heterogeneous, there is substantial variability in apparently equivalent samples. There are also differences in technique among the several laboratories and collecting agencies while at the same time there is a declared desire to ensure that results from the various programs are "compatible", "comparable" and "reliable". Presentation of a statistical evaluation is important in establishing confidence in the reported levels of contaminants. This is especially important in the Great Lakes region where the assessments will be used in judging successes under the Great Lakes Water Quality Agreement and the nature of additional, possibly costly control measures.

Levels and Precision of Contaminants in Rain and Air

Replicated results are available using a large (0.2 m^2) rain collector and from a prototype air sampler. The levels given here mainly indicate the levels of contaminants which can be expected in the several media and the present levels of imprecision. Sampling sites, during 1981-1984, are from across Canada -- Cree Lake is in northern Saskatchewan, Kouchibouguac is on the Atlantic coast in New Brunswick, Fifty Mile Point is on western Lake Ontario and the remaining locations are all around Lake Superior. These samples are singled out from among many others because triplicate samples were available for the sampling periods indicated. Other samples, however, indicate that the levels of the more prominent organochlorine contaminants are uniform (to within a factor of 2X to 3X) across Canada. For purposes of perspective, a concentration of 10 ng/L in rain falling on 80,000 km^2 (an area approximating Lake Superior) would result in a loading of nearly 1 metric tonne.

Analytes investigated with the rain sampler include the PCBs and most of the common organochlorine pesticides; those reported in Table 1 are those most commonly found from replicated (three) samplers (Strachan and Huneault, 1984; Strachan, 1985, 1988) placed to collect "ambient" samples rather than those from urban or known contamination sources. In all cases presented here, the results are for samples which were found in all three samplers but no additional discrimination has been applied.

The gas-air particle sampler (GAPS) is a prototype instrument designed to separate airborne particulates from vapour without filtration. It has only been evaluated for HCB and lindane (Lane et al., 1988) although there are some preliminary results with PCBs. The results in Table 2 are those from triplicate samples taken in a study at Little Turkey Lake north of Sault Ste. Marie (Johnson et al., 1989). Each sample was collected over approximately 46 hrs pumping air at 16.7 litres per minute; the "replicates" (three) were not exactly replicates since they were taken sequentially over a one week period.

Table 1: Environmental Means (and Relative Standard Deviations) of Rain Samples

Sample Description	Volume (litres)	α-HCH	γ-HCH	ΣPCBs	ΣDDT	Dield'n	HCB
		ng/L (percent)					
50MP, Apr. 1981	0.42			61(66)			
50MP, Apr. 1981	2.1	35(39)	7(20)	9(98)			
50MP, Apr. 1981	2.1	38(15)	11(15)	11(9)			
50MP, Apr. 1981	2.2	29(29)	10(28)	22(16)			
50MP, May 1981	1.0	10(31)	5(35)	32(45)			
TLake, June 1981	18.	5(3)	1(4)	1(36)		0.4(12)	
TLake June 1981	3.2	28(31)	7(60)	8(28)			
TLake, June 1981	4.2	48(41)	11(47)	3(20)			
TLake, June 1981	22.	19(39)	7(51)	1(65)	0.4(38)		
TLake, July 1981	15.	32(31)	9(14)	2(33)		0.6(12)	
IRoyale, June 1983	18.	65(21)	16(26)	16(29)	0.1(50)		0.1(60)
IRoyale, Aug. 1983	17.	40(15)	8(14)	5(62)	0.7(95)	1. (32)	0.01(100)
IRoyale, Setp. 1983	35.	19(11)	5(13)	1(53)	0.5(50)	0.5(27)	
Caribou, May 1983	7.9	9(16)	4(16)	48(21)	1. (53)	0.7(22)	
Caribou, June 1983	26.	46(3)	12(5)	6(56)	0.2(20)		0.1(12)
Caribou, Aug. 1983	34.	4(75)	2(27)	0.6(50)	0.2(100)	0.3(26)	0.02(69)
Caribou, Sept.1983	20.	6(3)	2(9)	2(75)	0.3(40)	0.3(25)	0.4(28)
Caribou, Oct. 1983	22.	5(16)	2(14)	2(51)	0.1(43)	0.3(6)	
Agawa, June 1984	21.	3(8)	2(17)	2(9)	0.2(22)	0.2(23)	0.03(24)
Agawa, July 1984	22.	4(11)	2(11)	5(70)	0.5(55)	0.6(37)	0.1(19)
Agawa, Aug. 1984	21.	3(24)	4(21)	0.1(41)	0.7(41)		
Agawa, Sept. 1984	26.	15(22)	5(14)	2(33)		0.9(33)	
Caribou, May 1984	57.	4(21)	2(26)	3(45)	2. (58)	0.5(55)	
Caribou, June 1984	19.	4(10)	2(16)	3(45)	0.6(40)	1.0(47)	0.03(13)
Caribou, July 1984	20.	7(7)	3(31)	3(21)	1.0(41)	0.9(28)	0.2(38)
Caribou, Aug. 1984	18.	4(22)	4(24)	4(56)	0.2(37)	1.3(17)	0.1(74)
Caribou, Sept.1984	20.	12(16)	4(7)	1(0)	0.7(10)	0.8(5)	
Cree, June 1984	21.	4(11)	4(12)	0.6(29)	0.1(45)		
Cree, July 1984	31.	0.9(22)	0.5(39)	3(4)	0.1(31)		0.01(16)
Cree, Sept. 1984	.4	53(3)	14(22)	17(14)	2. (29)		
Kou'ac, June 1984	43.	2(14)	1(14)		0.06(16)	0.07(46)	0.01(28)
Kou'ac, July 1984	50.	6(26)	3(40)	0.7(11)	0.6(55)		0.1(5)
Kou'ac, Aug. 1984	18.	1(47)	0.5(71)	2(104)	0.1(18)		
Kou'ac, Sept.1984	20.	66(53)	34(38)	7(46)	0.9(55)	2. (24)	0.1(28)
Mean Rel. Std. Dev. (%)		(22)	(25)	(40)	(43)	(27)	(37)

Table 2: Environmental Means (and Relative Standard Deviations) of Air Samples

Sample Description	α-HCH	γ-HCH	ΣPCBs	ΣDDT	Dield'n	HCB
	ng/m³ (percent)*					
TLake, Fall 1986	0.36(29)	0.06(47)				0.13(14)
TLake, Spring 1987	0.29(6)	0.10(16)				0.09(20)
TLake, Summer 1987	0.46(3)	0.04(11)				0.07(39)
TLake, Fall 1987	0.20(6)	0.03(0)				0.08(6)
Mean Rel. Std. Dev. (%)	(11)	(19)				(20)

* Each value represents the mean (and rel. std. dev.) of three separate samples of approximately 45 m³ all taken within one week.

There are no apparent systematic variations due to analyte level, sample size, location or season in the relative standard deviations in either the rain samples or in those for air. Mean levels for the imprecision estimates are therefore arrived at by simple averaging and are given in the two tables. It is apparent that the variabilities of the rain samples (Table 1) and those for the air (Table 2) are substantial and would affect the estimates of loadings required under the Agreement. When compared with the analytical variability for comparable levels of analytes, however, the contribution to the imprecision from the sampling process is seen to be moderate and may be the consequence of inherent variability rather than technique. In most cases where assessment of significance of the atmospheric input mechanism has been possible (Strachan and Eisenreich, 1988), even this level of variability does not negate the observation that this route is the dominant one for remote lakes and presumably also terrestrial areas.

The rain results are entirely from studies at NWRI and there is now building a body of information from the Ontario Ministry of the Environment and from the U. S. Environmental Protection Agency using the same resin field concentration and recovery methods. There are also data from a Great Lakes network which employs solvent extraction for concentration and in the very near future, there is to be a Great Lakes network utilizing the resin method for field processing of samples. It is partially with these different but similar systems in mind that a plea is made here for assistance in developing a quality assurance protocol now, before this network develops a "life" (i.e., direction) of its own!

The small amount of data available for the air samples pertains to the use of the prototype GAPS sampler in a limited number of situations. It too, however, may end up in the coming Great Lakes "International Atmospheric Deposition Network" called for under Annex 15 of the Great Lakes Water Agreement (1987 Protocol).

There will be a need for quality assurance in all aspects of this since there will be several collecting agencies involved, a number of laboratories and different analytical protocols. Much of the required guidance is indicated in the several reports of the different work groups at the last Ecological Quality Assurance Wokshop especially the practical aspects; those of us involved with establishing the atmospheric network look forward to hearing from those who have actively pursued the directions provided there.

References

Broddin, g., Cautreels, W. and van Cauwenberghe, K. 1980. On the aliphatic and polyaromatic hydrocarbon levels in urban and background aerosols from Belgium and the Netherlands. Atmos. Environ. 14: 895-901.

Grosjean, D. 1983. Polycyclic aromatic hydrocarbons in Los Angleles air from samples collected on Teflon, glass and quartz filters. Atmos. Environ. 17: 2565-2573.

International Joint Commission. 1987. Summary Report of the Workshop on Great Lakes Atmospheric Deposition (October 1986). Int. Joint Comm., Windsor, 41 p.

Konig, J., Funcke, W., Balfanz, E., Grosch, G. and Potts, F. 1980. Testing a high volume air sampler for quantitative collection of polycyclic aromatic hydrocarbons. Atmos. Environ. 14: 609-613.

Johnson, N. D., Lane, D. A., Schroeder, W. H., and Strachan, W. M. J. 1989. Measurements of selected organochlorine compounds in air near Ontario Lakes: Gas-particle relationships. Presented at 32nd Annual Conference, International Association Great Lakes Research, Madison, Wisc., June 1989.

Lane, D. A., Johnson, N. D., Barton, S. C., Thomas, G. H. S., and Schroeder, W. H. 1988. Development and evaluation of a novel gas and particle sampler for semivolatile chlorinated organic compounds in ambient air. Environ. Science & Tech. 22: 941-947.

Strachan, W. M. J. 1985. Organochlorine substances in the rainfall of Lake Superior: 1983. Environ. Toxic. & Chemistry 4: 677-683.

Strachan, W. M. J. 1988. Toxic contaminants in rainfall in Canada: 1984. Environ. Toxic. & Chemistry 7: 871-877.

Strachan, W. M. J. and Eisenreich, S. J. 1988. Mass Balancing of Toxic Chemicals in the Great Lakes: The Role of Atmospheric Deposition. Inter. Joint Comm., Windsor, 113 p.

Strachan, W. M. J. and Huneault, H. 1984. Automated rain sampler for trace organic substances. Environ. Science & Tech. 18: 127-130.

WORKGROUP SESSIONS

Group Reports

WORKGROUP 1: QUALITY ASSURANCE ISSUES RELATED TO LABORATORY ASSESSMENT OF AQUATIC ECOSYSTEMS

Participant List

Peter Chapman (Leader)
E.V.S. Consultants
195 Pemberton Ave.
North Vancouver, B.C. U7P 2R4
Phone (604) 986-4331
FAX (604) 662-8548

Brian Cumming
Queen's University
Biology Department
Kingston, Ontario
Canada K7L 3N6
Phone (613) 545-6193
FAX (613) 545-6617

James Lazorchak (Rapporteur)
U. S. Environmental Protection Agency
Aquatic Biology Program
Quality Assurance Research Division
Environmental Monitoring Systems Lab-Cinti
3411 Church St.
Cincinnati, Ohio 45244
Phone (513) 533-8114
FAX (513) 533-8181

Dan McKenzie
U.S. Environmental Protection Agency
Environmental Monitoring Assessment Program
Inland Surface Waters
Environmental Research Laboratory-Corvallis
200 SW 35th St.
Corvallis, Oregon 97333
Phone (503) 757-4666
FAX (503) 757-4335

Dennis McMullen
Technology Applications, Inc.
c/o U. S. Environmental Protection Agency
Environmental Monitoring Systems Lab-Cinti
3411 Church St.
Cincinnati, Ohio 45244
Phone (513) 533-8114
FAX (513) 533-8181

Dean Neptune
U.S. Environmental Protection Agency
Quality Assurance Management Staff
RD 680
401 M St. SW
Washington, D.C. 20460
Phone (202) 475-9464
FAX (202) 292-0929

Dave Peck
Lockheed-ESC
1050 E. Flamingo Rd.
Suite 209
Las Vegas, Nevada 89119

Phone (702) 734-3254
FAX (702) 796-1084

Martin Samoiloff
Bioquest International
7 Loyola Bay
Winnipeg, Manitoba R3T 3J7

Phone (204) 269-7264
FAX (204) 269-6897

Brian Schumacher
Lockheed-ESC
1050 E. Flamingo Rd.
Suite 209
Las Vegas, Nevada 89119

Phone (702) 734-3229
FAX (702) 796-1084

Mike Stainton
Fisheries and Oceans Canada
Freshwater Institute
501 University Crescent
Winnipeg, Manitoba R3T 2N6

Phone (204) 983-5176
FAX (204) 983-6285

QUALITY ASSURANCE ISSUES RELATED TO LABORATORY ASSESSMENT OF AQUATIC ECOSYSTEMS

Introduction

The workgroup began with participants introducing themselves and describing what they hoped to get out of the workgroup session. The workgroup leader then outlined major concepts, workgroup objectives, and discussion topics suggested for coverage in the session. Major issues discussed included the following:

- QA/QC in the context of legally defensible data vs. scientifically defensible data. Related to this was data collection for regulatory purposes vs. non-regulatory or non-compliance monitoring (e.g. ecological assessment monitoring) and data collection for private sector use (e.g. class action suits).

- A discussion of terminology, i.e. QA includes externally imposed requirements while QC consists of internally imposed requirements (self imposed). Some participants felt that QA should be imposed from outside entities in the regulatory setting while QA is imposed from inside as internal policy when data generation is for non-regulatory purposes.

- The following questions were asked: "What are the issues in toxicity tests?" and "What are we trying to improve in QA activities associated with toxicity tests?". Proposed issues included a) adjustment for control mortality (Abbott's formula), b) the need to improve reference toxicants, and c) the need to add more or improve existing test performance criteria (control mortality). An additional question arose concerning measures that should be considered to assess the quality of individuals performing toxicity tests.

- A discussion of the need for lab certification vs. use of methods with QA requirements that internally require essentially similar requirements as lab certification, i.e. performance criteria, control charts, assessment of precision, accuracy, and repeatability.

- A discussion concerning lab error vs. field error. Field error tends to be much greater than measurement error (in-lab).

- The use of a tiered approach involving multiple species instead of single-species testing was raised as an issue.

A list was then prepared by the leader on consensus issues that needed to be covered for the remainder of the session. These issues were:

- Certification/standardization/performance evaluations
    - How good are reported data?
    - How should quality of data be assessed? Performance measures.
- Replication for reliability/uncertainty levels.
    - Sources of error
    - Control adjustment
    - Reference toxicants
- Data uses - well defined
    - Regulatory v. non-regulatory use: does it make a difference?
    - Data presentation and reporting
- Present QA/QC requirements
    - What is acceptable?
    - What needs improvement? (Performance measures)
- Reporting of measurement error
    - What is it?
    - Should we only worry about overall error?
- Methods of Acceptance of new test systems (methods).

Workgroup Objectives

The workgroup leader presented an annotated outline of materials prepared before the workshop containing issues to be considered during this workgroup session as well as those proposed by the workshop organizers that morning. Issues included:

- Review and discussion, commensurate with participants experience and expertise, of techniques currently available to assure quality and representativeness of bioassay data (i.e. toxicity tests, laboratory and in-situ).

- Preparation of **generic** and **specific** decisions and recommendations.

- Coverage of the following major concepts:

  - PLANNING AND DESIGN
  - TRAINING
  - DATA COMPARABILITY
  - WHEN ARE DATA USEFUL (OR USELESS)

The workgroup leader briefed the workgroup on what he felt were important QA/QC concerns and issues for laboratory toxicity testing. (Planning and design = good science).

- Hypothesis Statements: Many reports and scientific journal papers have strayed away from hypothesis testing and have gone to more open ended investigations (e.g. let's go and collect information on this idea and see what we can interpret from the data).

- A PRIORI Data Handling Decisions: Good design and QA/QC considerations need to be prospective, not retrospective, in terms of statistical evaluation. Those individuals that have training in statistics or experimental design know that in hypothesis testing one needs to specify up front the significance level at which a hypothesis will be tested. This is very similar to the emphasis in the Data Quality Objective (DQO) process, i.e. What do you want to measure and what level of uncertainty are you willing to accept in the results?

- A PRIORI Evaluation of Needs: Things to consider in toxicity testing prior to beginning an investigation include:

  - Matching contaminants to methods
  - Selecting appropriate test organisms
  - Test conditions (e.g., exposure method(s))

- Estimation of Needed Resources

- A PRIORI Definition/Consideration of Possible Scenarios

- Sensitivity Analysis: As part of the hypothesis development process, some estimates of variance are needed. Pilot testing is useful in obtaining estimates of variance so that required sample sizes can be determined.

Toxicity Testing QA/QC Considerations

In considering QA/QC approaches to lab toxicity testing, the following data use considerations were outlined by the workgroup:

- Permit Compliance*
- Monitoring of Ecosystem Status/Condition*
  - Ecological Assessment Monitoring
  - Changes After Action
  - May Be: Short-Term
    Long-Term
    Site-Specific
- Superfund Site Assessment
- Class Action Litigation
- Screening-Level Assessments
- Research

* The only two areas of data use examined in detail during the workgroup session.

The workgroup felt that within the time available 2 areas of data use would be considered; permit compliance and regional and national scale ecological monitoring. Under each of these, the major issues to consider in designing and integrating QA/QC were project constraints, amount of replication, major sources of uncertainty, training, methods development, reporting error, and lab certification.

- Constraints Relevant to Permit Compliance Monitoring: These constraints are often the result of rules and regulations. Also, there is a tendency to select standardized methodology, resulting in data with little or no site relevance.

  All of the following apply to permit compliance and are administrative; those with an asterisk may also have components of decision-making.

  - Chain of Custody
  - Sample Holding Time/Conditions
  - Sample Representativeness*

- Spatial/temporal concerns
- Grab vs. Composite Sampling

- Method Conditions
  - Acute vs. Chronic Tests
  - Dilution Water
  - Photoperiod
  - Temperature
  - Aeration (yes/no/sometimes)
  - End-Points
  - Biological Variability (test stock)
- Data Summarization/Reporting*
  - False Positive/False Negative Results?
- Acceptable/Unacceptable Level of Uncertainty*
- Regulations (e.g. Permits)
- Resources
  - Includes Money and Available Methods

Constraints Relevant to Monitoring Ecosystem Status/Condition: All following apply to monitoring system condition. Differences from constraints to permit compliance are shown in bold print.

- Sample Collecting and Tracking (= Data Base):
  - **Not Chain of Custody.**
- Sample Holding Time/Conditions

- Sample Representativeness:
  - Spatial/Temporal Concerns
  - Grab vs. Composite Sampling
- Method Conditions:
  - Acute vs. Chronic Tests
  - Dilution Water
  - Photoperiod
  - Temperature
  - Aeration (yes/no/sometimes)
  - End-Points
  - Biological Variability (test stock)
- Data Summarization/Reporting
  - False Positives/False Negatives?
- Acceptable/Unacceptable Level of Uncertainty
- Regulations: **Not Applicable**
- Resources
  - Includes Money and Available Methods
- **Standardization**
  - **By Site (x time, i.e. years)**
  - **By Region (x time, i.e., years)**
  - **By Time**
  - **Data Comparability**
  - **What Decision Criteria Need to Change?**

- Replication: These consensus considerations apply to both permit compliance and monitoring of system condition:

  - How Much? How Many?

  - Dependent on Program (Science and Design) and Performance (Rules and Constraints) Requirements - Should Be Site Specific.

- Major Sources of Uncertainty: Consensus issues to consider in both areas of permit compliance and monitoring for system condition.

  - Test Variability:

    - Organism

    - End-Point

    - Method

    - Life-Stage(s)

    Use scientific judgement, not just regulation, to determine correctness and/or appropriateness.

  - Environmental Variability: Heterogeneous vs. Homogeneous Environments. Need to determine appropriate number of samples based on purpose.

  - Changes in Sample Before Testing: Ensure consistency in and appropriateness of sampling, holding, and handling of sample prior to testing.

  - Performance of Animals/Performance of Analyst: Use positive and negative controls, but when using reference toxicants be aware that one cannot separate variability due to animals from that due to the analyst, although reference toxicant tests implicitly include information on the sensitivity and quality of test animals.

- Training (Human Resources and Total Quality Management)

  - Educational Background

  - In-House Structure/Documentation

- Standard Operating Procedures
- Internal Training/Career Development
- External Training/Career Development
- Use of Performance Evaluations
- Need for Technical Training for Managers
- Need for Two-Way Communication

Methods Development:

This item refers to input/information provided by the methods supplier.

In the group's opinion, this item has not previously been clearly defined, and needs to be.

The following information was proposed as necessary for any methods development program:

- Optimum/Appropriate Laboratory Replication:
  - Need to evaluate method performance for a given matrix and a given end-point to allow informed choices to be made by user.
- Mode of Action of Toxicity Related to Recommended End-Points:
  - Allows for determination of appropriate and inappropriate reference toxicants.
  - Allows for determination of when (and when not) to use this test in relation to measured or suspected contaminants.
  - Allows interpretation of results of this test in combination with other tests (e.g. if this test does not react but others do, potential information is gained as to which toxicant(s) are bioavailable).
- Limitations to the Test:
  - Includes matrix interferences such as grain-size in sediment toxicity tests.

- Includes biotic interactions such as loading density effects (e.g. cannibalism in mysids).

- Acceptable Criteria for Controls
  - Especially important for partial or full life-cycle tests, e.g. not just percent mortality but also measures such as reproduction and growth.
- Minimum, Maximum, and Optimal Number of Organisms per Replicate:
- Sensitivity Analysis
  - What is the power of the test?

- Reporting Error
  - Report Performance-Based Measures
  - Report Error Components For Which Your Decision Rule is Most Sensitive:
    - Define error components
    - Determine which are most sensitive
    - Balance cost against information gained
    - Use best scientific judgement, not just rules

- Certification (= Performance Evaluation): Except for the first item below, these apply equally to Permit Compliance and Monitoring of Ecosystem Condition.
  - Necessary, but not Sufficient for Laboratories Performing Regulatory Testing.
  - Must be Certain, Defined Quality/Performance Conditions:
    - How this is done is up to each lab
  - Generic Only up to a Point; Often Program-Specific:
    - Controls (positive and negative)

- GLP: appropriate system acclimation
  sample tracking
  correct organisms (taxonomy)
  proper environmental quality conditions
  avoidance of contamination
  documentation

- Reference toxicant tests
  monthly for lab cultures
  each collection for field animals
  each batch for outside cultures

It was the consensus of the group that this type of evaluation should be performance-based. The group also said no to the idea of an aggressive umbrella agency/industry group responsible for laboratory certification.

Unresolved Issues

Note that these apply equally to Permit Compliance and Monitoring of Ecosystem Status/Condition.

- Reference Toxicants
  - Which one(s)?
  - Frequency?
- Acclimation for Early Life Stage Tests
- The "Right" Test for the "Right" Situation
- Performance Standards vs. Reference Toxicant Assays:
  - Standard Waters and Effluents
- "Non-Aggressive" Standards Organization
- Meshing "Good Science" with "Rules"
- Is Bias So Close to "Truth" that Precision = Accuracy?
  - Precision + Bias = Accuracy
  - Bias = Systematic Error

- "Truth" is Method-Specific; test organisms may show major differences

At this time, we can only approach a determination of bias by: standardization; use of reference toxicants; determination of coefficient of variation in laboratory round-robin and/or intercalibration testing

Summary

The group spent a good deal of time discussing lab certification. This issue appears to be of considerable importance to the environmental compliance program in both Canada and the United States. The group initially felt that some government regulation of this issue was necessary. However, it was concluded that, if good performance standards or criteria are developed for each method and results of performance are included with data reports, a self-governing system would work. Other issues that were of major concern to the group were training and reference toxicants or other toxicity testing performance criteria. There was a realization that present reference toxicant testing is insufficient because animal performance is not segregated from analyst performance. Training is imperative and must be emphasized more in QA plans. There is a need to establish a standardized protocol for methods development. Criteria should be set when developing methods. A consensus set of methods performance criteria could reduce the length of time spent by government entities in defending methods developed for use in the regulatory community. Hypothesis setting and defining objectives for a study are important considerations. QA in the absence of a hypothesis or objective is non-functional QA.

WORKGROUP 2: QUALITY ASSURANCE ISSUES RELATED TO FIELD ASSESSMENTS OF AQUATIC ECOSYSTEMS

Participant List

Mr. Michael T. Barbour (Leader)
EA Engineering, Science, and Technology, Inc.
15 Loveton Circle
Sparks, MD 21152
Phone (301) 771-4950
FAX (301) 771-4204

Mr. Gordon Craig
Beak Consultants Limited
14 Abacus Road
Brampton, ON L6T 5B7
Phone (416) 458-4044
FAX (416) 458-7303

Mr. Ian Davies
Freshwater Institute
Fisheries and Oceans Canada
501 University Cres.
Winnipeg, MB R3T 2N6
Phone (204) 983-5208
FAX (204) 983-6285

Ms. Kristin Day
Rivers Research Branch
National Water Research Institute
Environment Canada
867 Lakeshore Road
P.O. Box 5050
Burlington, ON L7R 4A6
Phone (416) 336-4659
FAX (416) 336-4989

Dr. Merete Johannessen
Norwegian Institute for Water Research (NIVA)
P.O. Box 69, Korsvoll
N-0808 Oslo 8
Norway
Phone (47) 2-235280
FAX (47) 2-394189

Mr. John Kingston
Laboratory Services Branch
Ontario Ministry of the Environment
125 Resources Road
P.O. Box 213
Rexdale, ON M9W 5L1
Phone (613) 235-5838
FAX (613) 545-6617

Dr. Dan Michael
Research Triangle Institute
1717 Massachusetts Ave., N.W.
Washington, D.C. 20036
Phone (202) 332-5102
FAX (202) 234-8761

Dr. John Paul
U.S. Environmental Protection Agency
27 Tarzwell Drive
Narragansett, RI 02882
Phone (401) 782-3037
FAX (401) 782-3030

Dr. Francis Philbert
Inland Waters Directorate - Ontario Region
Environment Canada
867 Lakeshore Road
P.O. Box 5050
Burlington, ON L7R 4A6
Phone (416) 336-4663
FAX (416) 336-4989

Dr. James Pollard
Lockheed Engineering & Sciences Co.
1050 E. Flamingo Road
Las Vegas, NV 89119
Phone (702) 734-3367
FAX (702) 796-1084

Mr. Stewart Thornley (Rapporteur)
Ontario Ministry of the Environment
985 Adelaide Street, South
London, ON N6E 1V3
Phone (519) 661-2200
FAX (519) 681-1011

Mr. Michael Walsh
Water Resources Branch - Great Lakes Section
Ontario Ministry of the Environment
135 St. Clair Avenue, West, Suite 100
Toronto, ON M4V 1P5
Phone (416) 323-4946
FAX (416) 965-9807

Mr. Don Wismer
Environmental Studies and Assessment Dept.
Ontario Hydro
700 University Avenue, Rm H10 E4
Toronto, ON M5G 1X6
Phone (416) 592-8596
FAX (416) 978-0156

QUALITY ASSURANCE ISSUES RELATED TO FIELD ASSESSMENT OF AQUATIC ECOSYSTEMS

Introduction

Our basic objective was to develop a generic quality assurance approach for an ecological study of an aquatic system. We identified the major activities (i.e., sampling design, field operations, laboratory activities, data analysis, reporting) required for an ecological survey of an aquatic system and defined the related quality concerns needed to control the project activities. The generic concept of quality assurance was applied to a fabricated case study to illustrate the process. The data quality objective (DQO) process is outlined in the first paper of these Proceedings, and provides a logical sequence of six steps required to lay out the foundation of an appropriate research study plan (Figure 1). The major activities identified under our generic project and case study were discussed in the context of the DQO logic flow sequence.

Statement of Problem

To begin the process, a problem statement was developed in the form of a hypothetical question: "Based on information from a study, what action is required by a discharger to protect use(s)?". This general question required that we define the particular situation in the case study. The hypothetical case study consisted of an industry discharging compound "X" into a large river system. The purpose of the study was to ascertain whether biological impairment, according to the designated use of the receiving river, was present and attributable to the discharger in question.

In analyzing the problem statement, four major components were identified (Table 1). First, the problem should be related to the decision process. It is not sufficient to simply determine impairment; the consequence must also be stated. In reference to our case study, "the action required by the discharger" was the ultimate question. The second component was the broader application of the study concept and hence the need to relate to uses. In the case study, three uses were defined to provide examples for the problem statement. The third component involved the "users" in the decision process. The purpose of this involvement is to communicate the intent of the study and desired question to be answered prior to implementation. The "users" in the case study included a regulatory agency, the discharger, and the public. The fourth component to the problem statement is to identify specific questions that need to be answered. The development of specific questions for the case study required that the effects of compound "X" be delineated. Therefore, three effects were constructed which stated that compound "X" (1) is toxic to mayflies, (2) bioaccumulates in trout, and (3) has a high affinity for adsorption to sediment. From these effects specific questions could be asked.

In analyzing the case study from our given field perspective, our challenge was to verify an existing problem (i.e., are uses impaired?), and to determine the probable cause (i.e., what are the contaminants?). To ascertain impairment, the status of the biota and habitat as reflected by reference conditions must be examined. In addition, temporal and spatial differences are important factors in assessing impairment. The entire DQO process is iterative, and throughout these initial steps it was necessary to repeatedly refine the initial question as more information on the case study was made available.

Study of Variables and the Logic Statement

The next step in the DQO logic process was to identify all variables which may affect the problem (Table 2). The variables were classified as abiotic or biotic. Once all possible variables are identified, those that are pertinent to addressing the specific question are selected. For instance, the first specific question of the case study is as follows: Is the mayfly population impaired? A few abiotic and biotic variables were selected as an example in our case study (Table 3).

From selected variables, logic statements are developed. A logic statement can be partitioned into three parts for each variable: a result of the processed data, a domain for judgement of effects, and the logic or decision criteria (Table 4). The example variable for this case study was the presence or absence of mayflies. For purposes of this example, the absence of mayflies denoted adverse effects. The result to be evaluated was the percentage of units (samples) within the areas of interest (stations). The occurrence of effects would be judged from a comparison of treatment (affected) and background (natural variation) conditions. Two domain alternatives for the case study were established: (1) distance intervals, or reaches, could be delineated (e.g., 2 X mixing zone area); or (2) the concentration levels of compound "X" in the sediments could be measured and used to establish the domain. The logic or decision criterion then might be the percent of units without mayflies in the treatment areas compared to that of the background area. As an example, if the number of samples or units in the treatment area without mayflies exceeded by as much as 50 percent the number of units in the reference area with no mayflies, then an impairment would be determined.

Level of Uncertainty

The level of uncertainty associated with the logic statement must be established to provide a level of confidence in the decision. Often, the resources or cost investment may limit control over the level of uncertainty which can be achieved. An analysis is required of the error and the consequences of that error to influence that decision. There are two possiole error scenarios, as illustrated in the following example. (1) A false negative (ß) error results when there is no detected difference between the treatment and reference areas in estimating mayfly abundance. However, the truth is that there is a difference whereby the mayfly abundance is lower in the treatment area.

The consequence of this error (false negative) is that an ecological impact goes undetected. In this scenario, the population of mayflies has declined, with the probability that the fishery also would decline. The implications of this error may be to repeat the sampling scenario at an associated additional cost. Other implications may be a loss of an important fishery and loss of credibility with the public. (2) The second error scenario is the false positive (α), which states that a problem (difference) is deduced from the results when, in truth, there is no difference between the treatment and the reference area. The consequence of this error is that the discharger may be required to impose unnecessary mitigation when compound "X" is in fact not affecting the mayfly population in the treatment area. Implications in this error scenario are the cost of unnecessary treatment and a possible loss of credibility for the researcher with the industry. The limits on these two types of uncertainties should be established so the researcher can design a program that meets the DQO's laid out in the study design. The quality control samples taken during the implementation phase will provide the data that are needed to determine whether the DQO's have been met.

Quality Control Elements

To control error, it is necessary to identify the quality control (QC) elements associated with each study activity. Five activities are associated with ecological studies:

. Sampling Design
. Field Operations
. Laboratory Activities
. Data Analysis
. Reporting

It is possible to identify potential error sources for each activity and the respective quality control elements needed to control those errors. From this information a matrix can be constructed showing the interaction between the QC element and the potential error.

Sampling Design -- The primary question of the sampling design is whether the data collected as per the sampling design address the central question. The QC elements need to assure that this is the case. The QC elements include conducting a pilot study, incorporating environmental strata, incorporating historical data, including replicates, and choosing appropriate equipment (Table 5). Several considerations relating to potential error in the sampling design can be described. The resources available may be too limiting to adequately address the problem. Limited resources might be due to either developing too broad a study or inadequate expertise and facilities. Logistical problems may arise which will hinder the successful completion of the study. The response variables selected for the study may be inappropriate or insufficient to address the question. Weather contingencies and seasonality issues are two other considerations. The selection of site locations is critical in determining impairment or the status of conditions. Both habitat and population variability are

important in defining background conditions before impairment can be assuredly detected. The choice of equipment is a consideration for the proper design of collecting pertinent data. More than one QC element may be instrumental in reducing potential error for each consideration. This multiple interaction between QC elements and considerations of potential error strengthens the overall quality of the program. Invariably, one QC element will become more important than others in influencing several areas of potential error. For sampling design, the conduct of a pilot study or reconnaissance is probably the single most important element. However, the availability of historical data greatly enhances the design of the sampling plan.

Field Operations -- For the field operations aspect of an ecological study, the major QC elements are instrument calibration and maintenance, crew training and evaluation, field equipment, sample handling, and additional effort checks (Table 6). The potential errors in field operations can be classified into eight general categories. One of the potential error sources is climate. Successful field operations hinge on adequate preparation for adverse climatological conditions. Proper location of the sample sites is considered an important potential error that can be minimized using the appropriate equipment (which has been properly maintained), and a well-trained field crew. Both sampling equipment inefficiency and human (field crew) inefficiency are potential errors that can quickly invalidate the collection efforts if not controlled. Field notes are integral to the documentation of activities and can be a potential error source. Other considerations for potential errors are sample processing, transportation, and tracking. It was found that training may be one of the most important QC elements with application to most of the particular considerations for potential error.

Laboratory Operations -- The QC elements of Laboratory Operations were classified as sorting and verification, taxonomy, duplicate processing, archival procedures, training, and data handling (Table 7). Potential error sources associated with sample handling are sample tracking, im proper storage, and sample preparation. All of these are best controlled by staff training. Other important QC elements necessary for controlling sample handling errors are sorting and verification procedures and archival procedures. Controlling taxonomic error requires well-trained staff and available expertise to verify identifications. Counting error and sorting efficiency are usually the most prominent error considerations and may be controlled via training, duplicate processing, sorting, and verification procedures. Laboratory and computer data records are two very important potential error sources that, through data handling and staff training, can be minimized.

Data Analysis -- Training and data handling and reporting are important QC elements for the data analysis phase (Table 8). Standardized databases and analysis programs were considered important to control programming errors and misinterpretation of data. Peer review and range control of values are the last two QC elements for data analysis. Potential errors can occur if inappropriate statistics are used to analyze the data. In this case, erroneous conclusions can be drawn regarding trends in the data. Undetected errors in the database or programming can be disastrous to an

appropriate interpretation. Problems in managing the database can occur if steps aren't taken to oversee the data handling, analysis, and summarization. A final consideration is the possible misinterpretation of the findings. All of the potential errors are best controlled by adequately training the staff. The uses of standard database and analysis computer software packages provide ready tools that have already been subjected to quality assurance programs.

Reporting -- QC elements include training, peer review, the use of a technical editor and standard format (Table 9). Transcription error and a poor presentation can occur if care is not taken to provide adequate training and appropriate review. The use of obscure language can often mislead the reader. Peer review and the review by a technical editor can be important deterrents to this potential error. Whether the central question was addressed by this study can be an error in reporting. The report should strive to answer the main question, and if not, the reporting format is faulty.

Summary

All of these aspects of a quality assurance program that form the basis of the DQO process are inherently interrelated so that the questions and procedures are refined as the study design is detailed. The interrelationship of the DQO process is not directly a linear process as indicated at the beginning of this report, but is, in fact, an iterative process as illustrated in Figure 2. The initial statement of the problem can evolve as specific questions are identified and existing data are reviewed. This circular refinement is necessary to prepare the appropriate study design of the project (Portion I of Figure 2). The second iterative step in the DQO process (Portion II of Figure 2) is the identification and selection of the variables to be measured as they influence the refinement of the specific questions. Once the specific questions and pertinent variables are finalized in the study plan, then the logic statement for each variable can be developed, acceptable uncertainty established, and sources of potential error identified. The proper QC elements are incorporated to control and minimize the error sources with the ultimate objective of assuring the acceptability of the study and addressing the central question or problem of concern.

Future Considerations

The future considerations discussed by Workgroup 2 for Quality Assurance of ecological studies were to define to a greater extent the QA/QC control points where error can be identified and quantified. Five major considerations are as follows:

. Termination Control Point
. Field Crew Efficiency
. Additional Performance Effect Criteria
. Statistical Power Analysis
. Interlaboratory Training and Calibration

A termination control point may be necessary to end the program if the study design is inadequate to address the problem or the problem is too ambitious for the available resources. What sort of decision matrix can be established to provide a controlled decision for project termination? Field crew efficiency is integral to the study and control points should be established for evaluating this efficiency and deciding whether more training is necessary before the study proceeds any further. Additional effort, performance criteria and statistical power analyses are not new concepts to quality assurance, but a delineation of how these aspects are integrated into the project to enhance the integrity of the data needs to be considered. Interlaboratory training and calibration are commonplace to analytical and toxicity testing programs. This concept needs to be explored for application to ecological studies.

TABLE 1: STATEMENT OF PROBLEM

Generic Question: Based on information from a study, what action is required by a discharger to protect use(s)?

GENERIC	CASE STUDY
. Relate problem to the decision process	. Action required by discharger
. Broader application to the concept of the study	. Uses: 1) Sustainable species fishery 2) Recreational uses 3) Aesthetics
. Involve users in the decision process	. Regulatory agency, public, discharger
. Identify specific questions that need answering	. 1) Compound X is toxic to mayflies 2) Compound X bioaccumulates in trout 3) Compound X has a high affinity for sediment adsorption

TABLE 2: IDENTIFY ALL VARIABLES

ABIOTIC	BIOTIC
Discharge Constituents	Species Abundance
Discharge Loadings	Species Composition
Discharge Dispersion	Species Trophic Status
Outfall Length	Presence/Absence of Indicators
Compound X - Outfall	Life Cycle Analysis
Compound X - Receiving Water	Species Biomass
Receiving Water Characteristics	Growth Rates
Water Clarity	Compound X in Biota (Body Burden)
Substrate	In-situ Bioassay
Water Depth	

TABLE 3: SELECT VARIABLES TO BE MEASURED

ABIOTIC	BIOTIC
. Concentration of compound X in surficial sediments	. Presence/absence of mayflies
. Concentration of compound X in water	. Mayfly composition and distribution
. Concentration of compound X in outfall	. Relative abundance of mayflies
. Characterization of substrate	. In-situ mayfly bioassay

TABLE 4: DEVELOP LOGIC STATEMENT

VARIABLE	RESULTS	DOMAIN	LOGIC
Presence/ absence of mayflies	Percent of units within areas of interest without mayflies	Treatment (affected) vs. background (natural variation) 1. Distance intervals (reaches) (e.g., 2 X mixing zone area) 2. Concentration levels of compound 'X' in sediments	If percent W/O mayflies in treated areas is X% (e.g. 50%) more than found in background - take action - i.e., if X_2-X_1 > 50%, where X_1 = number of units (sq. m. samples) W/O mayflies in reference area, and X_2 = number of units W/O mayflies in treatment area

TABLE 5: SAMPLING DESIGN

Consideration (Potential Error)	Pilot Study	Environmental Strata	Historical Data	Replicates	Choice
RESOURCES AVAILABLE	X	X		X	
LOGISTICS	X				X
RESPONSE VARIABLES	X		X		
WEATHER		X	X		
SEASONALITY	X	X	X		
SITE LOCATION	X	X	X		
HABITAT VARIABILITY	X	X	X	X	
POPULATION VARIABILITY	X		X	X	X
EQUIPMENT	X				X

TABLE 6: FIELD OPERATION

Consideration (Potential Error)	QC Element: Instrument Calibration & Maintenance	Crew Training & Evaluation	Field Equipment	Sample Handling	Additional Efforts Checks
CLIMATE		X	X		
SITE LOCATION	X	X	X		
SAMPLING EQUIPMENT EFFICIENCY	X	X	X		
HUMAN (EQUIPMENT USE)	X	X			X
FIELD NOTES		X		X	
SAMPLES PROCESSING	X	X	X	X	
SAMPLE TRANSPORTATION	X	X	X	X	
SAMPLE TRACKING	X	X	X	X	

TABLE 7: LABORATORY

Consideration (Potential Error)	QC Elements: Sorting & Verification	Taxonomy	Duplic. Process	Archives	Training	Data Handling
SAMPLING TRACKING	X			X	X	X
IMPROPER STORAGE				X	X	
SAMPLE PREPARATION	X		X	X		
REFERENCE ERROR (TAXONOMY)		X			X	
TAXONOMIC ERROR (HUMAN)		X	X	X	X	
COUNTING ERROR	X		X		X	
SORTING EFFICIENCY	X		X		X	
DATA RECORDS (LAB)	X	X	X		X	X
DATA RECORDS (COMPUTER)					X	X

TABLE 8: ANALYSIS -- QC ELEMENTS AND POTENTIAL ERROR

Consideration (Potential Error)	QC Elements					
	Training	Data Handling & Reporting	Standardized Database	Standardized Analysis	Peer Review	Range Control
INAPPROPRIATE STATISTICS	X				X	
ERRORS IN DATABASE	X	X	X			X
DATABASE MANAGEMENT	X	X	X			
PROGRAMMING ERRORS	X	X	X	X		
MISINTERPRET ANALYSIS	X			X	X	

TABLE 9: REPORTING

Consideration (Potential Error)	QC Elements			
	Training	Peer Review	Technical Editor	Standard Format
TRANSCRIPTION ERROR	X			
POOR PRESENTATION	X	X	X	X
OBSCURE LANGUAGE	X	X	X	X
DID WE ADDRESS QUESTION	X	X		

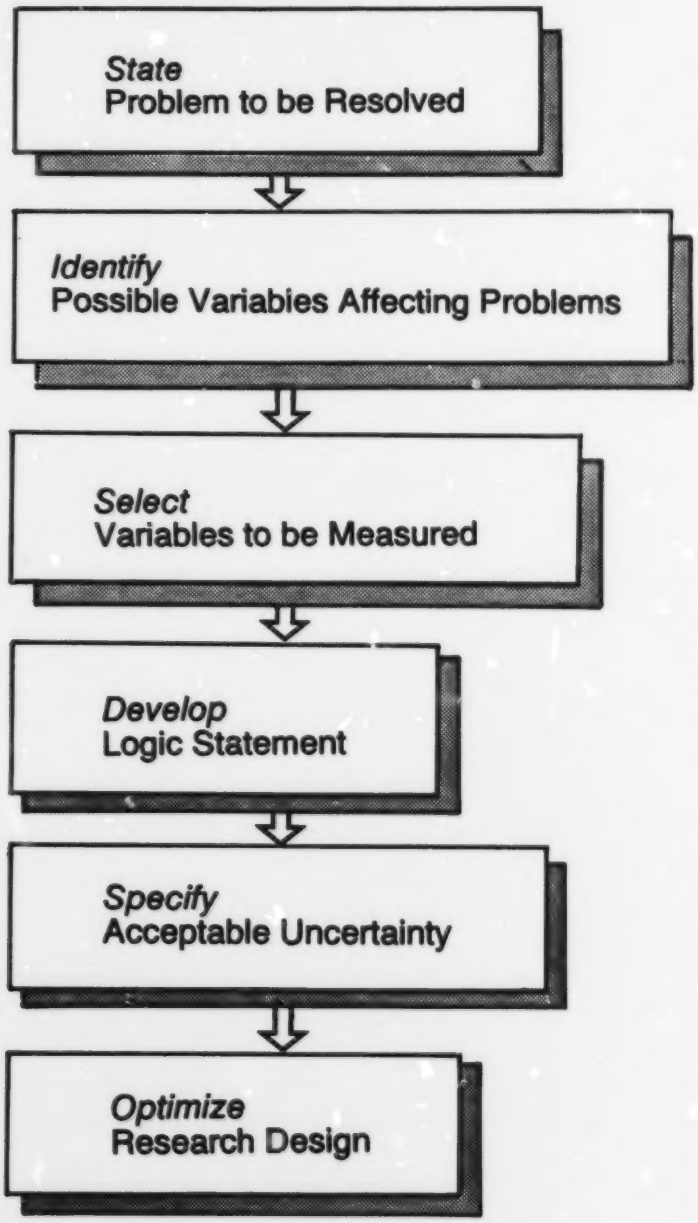


Figure 1. Sequential logic of Data Quality Objective (DQO) Research.

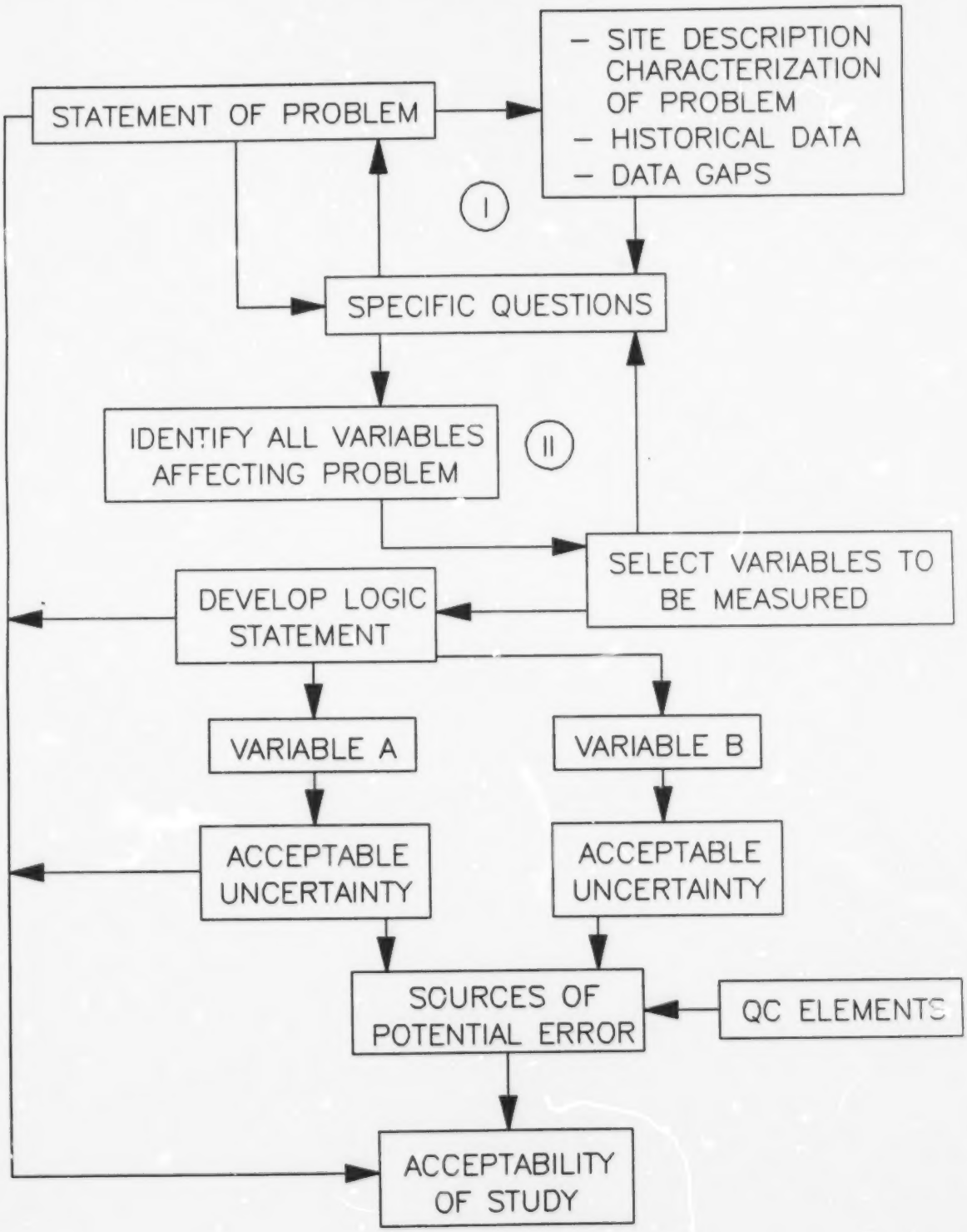


FIGURE 2. Summary of DQO Process for Ecological Studies.

WORKGROUP 3: QUALITY ASSURANCE ISSUES ASSOCIATED WITH TERRESTRIAL ECOSYSTEM MONITORING

Participant List

Bob Breckenridge
Idaho National Engineering Laboratory
P.O. Box 1625-2213
Idaho Falls, ID 83402
Phone (208) 526-0757
FAX (208) 526-2933

William Burkman (Rapporteur)
NSI Tech. Services Corp.
c/o USDA Forest Service, NEFES
5 Radnor Corporate Center, Suite 200
100 Matsonford Road
Radnor, PA 19087
Phone (215) 975-4076
FAX (215) 975-4154

Alain Dufresne
Macdonald College, McGill University
21111 Lakeshore Road
Ste. Anne de Bellevue, PQ H9X 1CO
Phone (514) 398-7942
FAX (514) 398-7983

Kate Dwire
NSI Tech. Services Corp.
1600 SW Western Boulevard
Corvallis, OR 97333
Phone (503) 757-4666
FAX (503) 757-4335

John Innes
Forestry Commission
Alice Holt Lodge, Wrecclesham
Farnham, Surrey GU10 4LH
England
Phone 420-22255
FAX 420-23653

Craig Kinch
Beak Consultants Ltd.
14 Abacus Road
Brampton, ON L6T 3B6
Phone (416) 458-4044
FAX (416) 458-7303

Linda Kirkland
Office of Research and Development (RD-680)
EPA/OMMSQA
401 M Street SW
Washington, DC 20460
Phone (202) 382-5775
FAX (202) 252-0929

Denis Laflamme
Dir. Recherche et Development (MER)
Ministere de l'Energie et des Resources
2700 Rue Einstein
Ste.-Foy, PQ G1P 3W8
Phone (418) 643-7994
FAX (418) 643-2165

Garry Linsey
Freshwater Institute
Fisheries and Oceans Canada
501 University Cres.
Winnipeg, MB R3T 2N6
Phone (204) 983-5177
FAX (204) 983-6285

Marius Marsh
Ontario Ministry of the Environment
Air Resources Branch
Phytotoxicology Section
880 Bay Street
Toronto, ON M5S 1Z8
Phone (416) 456-2505
FAX (416) 456-2246

Imants Millers
USDA Forest Service, NA State and Private Forestry
P.O. Box 640
Durham, NH 03842
Phone (603) 868-5719
FAX (603) 868-1538

Ian Morrison (Leader)
Forestry Canada
Great Lakes Forestry Centre
1219 Queen Street East
P.O. Box 640
Sault Ste. Marie, ON P6A 5M7
Phone (705) 949-9461
FAX (705) 759-5700

Rick Van Remortel
Lockheed Engineering & Sciences Co.
1050 E. Flamingo Road, Suite 200
Las Vegas, NV 89119
Phone (702) 734-3295
FAX (702) 796-1084

QUALITY ASSURANCE ISSUES ASSOCIATED WITH TERRESTRIAL ECOSYSTEM MONITORING

Workgroup Objective

The objective of this workgroup was to develop a general approach for selecting methods for a terrestrial ecosystem monitoring program. A forestry or vegetation monitoring program was used as the model. The focus could have easily been on data management, statistical design, or any other aspect of program planning. The approach began with the statement of the problem, proceeded through the decision process, to data collection and reporting. The seven steps are:

1. Question Definition
2. Proposed Variable Prioritization
3. Method Selection
4. Method Documentation
5. Training
6. Data Evaluation and Validation
7. Use of Data Quality Information

The procedure followed was to develop general guidelines, areas of consideration, identification of potential problems, and attributes for each of the seven steps.

Question Definition

It was assumed by the workgroup that higher level management or the political establishment proposed a question of concern, e.g. "Why is the supply of lumber in eastern North America so low?". This question implies that tree growth is reduced and/or mortality is increased. The first step is to recognize and state the potential constraints in answering this question. These constraints may be: available funding, personnel limits, temporal and/or spatial considerations. Sample design and plot layout should also be recognized and included in this stage. The specific area of interest is then determined and the testable hypotheses are stated. This process should develop a clear, concise, potentially-achievable objective to plan the monitoring program.

Proposed Variable Prioritization

After the statement of the problem and the development of clear objectives, the variables which will be assessed in order to achieve the objectives are listed and prioritized. Generally with ecosystem monitoring, a staff of individuals is assembled from various research areas to develop the research program. When working on a team format, each individual has a set of "personal preference variables" and includes these

on the list. The balance of necessary variables and optional variables needs to be considered and a consensus on variables selected is reached.

An integral step in this process is to understand and define the population of interest. This step includes determining the expected range and variation of the variables of interest. Historical or published information should be used whenever possible and the importance of this information should be recognized. When this type of information is not available, a pilot or scoping study is needed. The pilot study should also be used to understand and identify the sources of measurement error in addition to quantifying the normal variation of the variables.

Method Selection

The first step is to determine and state the requirements of the data quality parameters. These parameters are principally accuracy and precision, although other parameters such as completeness, representativeness, and comparability are also used. Different procedures exist for data collection for a particular variable and often the data quality will differ depending on the method. Thus, it is important to understand all of the potential methods and the inherent quality of the data that each produces.

The second step is to determine the cost of a data point arrived at by use of a particular method. Extremely detailed, involved procedures generally have a high data quality but the cost for each data point may also be very high. The selection of a method will involve the balancing of data quality and cost within the framework of the specific objectives and sampling design.

The various sources of measurement error should be stated and understood. Often this information can be retrieved from existing data or determined from the pilot study.

Method Documentation

After the selection of the method it is extremely important to develop detailed method documentation. A complete method description should include materials and equipment required, procedures for each step, the performance criteria (data quality), procedures that will be used to evaluate performance, expected data ranges, and data codes.

A large monitoring program may include many variables, resulting in a bulky methods manual. It may be necessary to develop a briefer version for every-day use. This may be a one page card or a pocket-sized manual, written in clear, concise language. The complete manual should be available to crews as a reference and is very useful for documentation of data collection procedures for future use, especially if the project data are used by subsequent groups. The complete manual should provide the necessary documentation for data users to determine if the procedures are appropriate.

It is important that the manual describe as many as possible potential interpretations and/or situations. This should increase the comparability between individuals. During the early years of a program the manual may have to be revised after each field season to clarify interpretation problems identified during the previous field season.

Training

Effective training is important to ensure data quality and comparability. In designing a training session, it is important to recognize the education and experience of the participants so the training session can be designed to maximize the areas covered. Some methods may require specific backgrounds and it may be necessary to establish minimum education and/or experience standards. All project personnel including field crew supervisors, data management personnel, and program management, should also participate in the training session.

The site used for training should also encompass the expected range of natural variation on the plots. Particular situations to include during the training session can be determined by recognizing the expected or normal variation of the variable in the population of interest.

The training session should be structured in discrete units, possibly by variable or groups of variables. By using discrete units, the instructor with expertise with a particular variable or set of variables can be selected to conduct training for that unit. Evaluations of personnel performance in discrete units provides immediate feedback on the effectiveness of the training. In addition, this procedure identifies individuals who may need more intensive training for particular areas of data collection prior to actual data collection, or problem areas in method documentation.

A complete evaluation or test of the entire set of variables should be conducted prior to collection of actual data. This may be structured under simulated field conditions and can estimate the accuracy of the method prior to data collection. The comparability among individuals can be determined from all crew personnel measuring the same plots and/or trees. With a large number of individuals, the accuracy of a method can be determined because the "true" value can be estimated from the testing results. The personnel trained should also evaluate individual instructors and the effectiveness of the training session.

Data Evaluation and Validation

The performance evaluation procedures established in the method documentation should be implemented during the data collection phase of the project. It was generally accepted that a repeat measurement of a subset of plots be undertaken to determine acceptance based on performance criteria.

An unresolved issue was how to utilize the results of repeat measurements. If the repeat measurement value was considered to be the "true" value, then accuracy and precision could be estimated. If the "true" value concept was not accepted, then only precision could be estimated. Methods used to determine the "true" value could be a reference crew, standard reference plots, or standardized photos. There are different problems associated with each of these approaches. The underlying problem involves the willingness to accept the reference crew, plot, or photo as the "true" value.

The plots selected for repeat measurement should be determined at random. This ensures that the field crews will measure each plot carefully because each plot has an equal chance of being remeasured. This is comparable to the blind reference samples used in chemical analyses.

It was also stressed that reviews (audits) of the field crews be used to ensure the implementation of methodology. Field reviews generally provide qualitative instead of quantitative information, but this information can be useful. The field reviews can also be used to summarize field crew concerns with the methods.

Another issue raised dealt with poor quality data determined from repeat measurements. The use of training sessions and good documentation should decrease the amount of poor quality data. If a crew did have problems then that crew should be immediately retrained. The second question is what to do with the data collected through the period in question. If time and/or money allows, the plots should be remeasured. If not, the data should be flagged as questionable.

The use of data recorders was also discussed. The use of data recorders eliminates many of the problems associated with the recording of data on tally sheets, such as transcription errors associated with entry into computer files or the omission of data. Data recorders can include range checks, code checks, and validation checks. Tally sheet backups should always be available.

Use of Data Quality Information

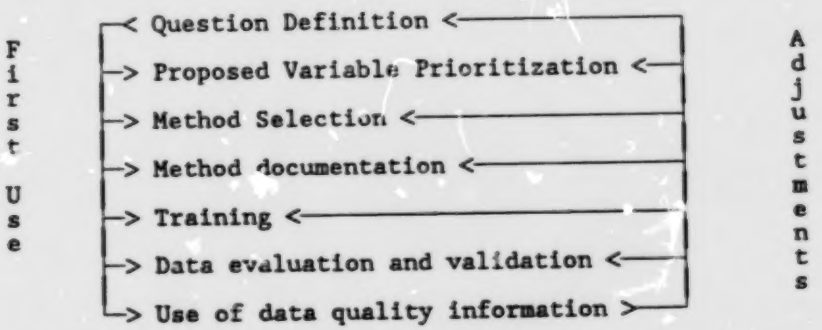
The data quality information should always be attached to the results rather than remain as a separate report or document. The summary should be kept simple without sacrificing information. An unresolved issue is whether to standardize the statistics used for the summary of data quality results. The attachment of the data quality information with the program results allows the reader to interpret the conclusions and the appropriateness of the selected method in the context of data quality. The data quality information can also be used to quantify the method data quality, especially if the performance criteria were estimated from historical data.

Qualitative information can also be useful in interpreting the results or in evaluating the effectiveness of a method. This information can be summarized from training session evaluations, field review reports, and/or end-of-field-season crew

debriefings. This information can be used to identify interpretation problems with method documentation and/or alter the emphasis of the training session for subsequent field seasons.

Relationship Between the Various Steps

During the first use of this process the steps are generally followed in a stepwise progression. Each of the steps in this process is related to all of the other steps. The program team needs to have constant communication and coordination because all of the steps are generally occurring simultaneously. If changes are made in any of the steps, such as a change in method, it may be necessary to make changes in other steps. In subsequent years, data quality information may be used to modify the question, change the method, improve the training session, or other uses. It is then necessary to revise the affected information and readjust the program. This process should be viewed as continually changing to improve the overall program results. The flow chart is shown below.


First Use
< Question Definition <
-> Proposed Variable Prioritization <
-> Method Selection <
-> Method documentation <
-> Training <
-> Data evaluation and validation <
-> Use of data quality information >
Adjustments

Conclusions

Data quality should be planned for and organized during the development of a program. To accomplish this, QA personnel need to be involved with overall project personnel, not as a separate group. This also applies to statistical and data management personnel. These steps provide a logical progression from problem statement to collection of appropriate data to make a conclusion. This process doesn't cover all of the issues and should not be viewed as complete, because often different situations or common sense will dictate changes in the process.

Unresolved Issues

1. How to assess accuracy of field measurements.

2. How to deal correctly with questionable data.

3. How to assess comparability between years. This is a major concern when the focus of the project temporal change.

4. How to keep field crew personnel involved with and enthusiastic about the project. As the crew loses interest, data quality will suffer.

5. Whether to standardize data quality statistics.

6. How to remove the separation between project management and QA. What is the best process to combine these aspects?

WORKGROUP 4: QUALITY ASSURANCE RELATED TO CONTROLLED FIELD AND LABORATORY EXPERIMENTS IN TERRESTRIAL ECOSYSTEMS

Participant List

Robert Mickler (Leader)
NSI Technology Servcies Inc.
c/o U.S. Forest Servcie
3041 Cornwallis Road
Research Triangle Park, N.C. 27709
Phone (919) 549-4022 (Ext. 4928)
FAX (919) 549-4047

Dave Wotton (Rapporteur)
Manitoba Environment
Bldg. 2, 139 Tuxedo Avenue
Winnipeg, Manitoba R3N 0H6
Phone (204) 945-7081
FAX (204) 945-5229

Donald Vereault II
Sawyer Environmental Center
WIMP Project
University of Maine, Orono
Orono, Maine
Phone (207) 581-3289
FAX (207) 581-1604

Alain LePage
Forestry Canada
Region of Quebec
1055 P.F.D.S., Sainte-Foy
Quebec, G1V 4C7
Phone (418) 648-2582
FAX (418) 648-5849

James. E. Gaskin
Water Quality Branch
Environment Canada
Ottawa-Hull, Ontario
Phone (819) 953-3117
FAX (819) 997-8701

Hélène Lalonde
Dept. of Renewable Resources
MacDonald College
21, 111 Lakeshore
St. Anne de Bellevue, Quebec H9X 1C0
Phone (514) 398-7942
FAX (514) 398-7083

Lisa J. Ellingson
NSI Technology Services Corp.
200 SW 35th Street
Corvallis, OR 97333
Phone (503) 757-4666
FAX (503) 757-4335

Steve Cline
NSI Technology Services Corporation
200 SW 35th Street
Corvallis, OR 97333

Phone (503) 757-4666
FAX (503) 757-4335

Craig Palmer
NSI Technology Services Corporation
200 SW 35th Street
Corvallis, OR 97333

Phone (503) 757-4666
FAX (503) 757-4335

QUALITY ASSURANCE RELATED TO CONTROLLED FIELD AND LABORATORY EXPERIMENTS IN TERRESTRIAL ECOSYSTEMS

Introduction and Objectives

The nine individuals participating in Workgroup 4 were from a diverse background of research interests and experiences. As a result discussions were stimulating and productive in a broad spectrum of terrestrial research. Early in deliberations the group elected to use a current research project as a case study.

A study to evaluate quantitative fertilization as an assesment tool for forest decline in the Montmorency Forests of Quebec was selected. The study incorporated three principal components of terrestrial research. Field sampling using a lysimeter collection technique, laboratory analysis to determine background chemistry of soils, and a controlled environment study in growth chambers to evaluate tree seedling response.

In brief, soils from the Montmorency Forest were sampled from selected sites and characterized for development of soil solutions used as treatments in growth chamber experiments. Sugar maple and red spruce seedlings were to be evaluated following treatments consisting of varied levels of fertility, aluminum content, CO_2 content and acid fog. A description of the case study and results of discussion are summarized in the latter half of the paper.

Priority tasks were established to address the quality assurance requirements of the case study. Three principal tasks were identified:

1. Establish a Quality Assurance Plan

   - Assemble research team
   - Document procedures
   - Develop a training program
   - Develop a reporting format

2. Determine what quality assurance and quality control information is required.

   - Ensure study design will achieve objectives
   - Develop critical questions and answers for each phase
   - Determine the QA indicators for each component

3. Develop a Data Management Plan

   - Determine who is the final audience
   - Determine what information is required
   - Develop a reporting schedule

Develop a reporting format for QA and results

QA Requirements

The Quality Assurance Plan was deemed the most critical component of effective research design. As a result the workgroup concentrated on this component. Key to the QA plan is the assembly of the research team. This should include all participants in the project from management to technical support. It was well recognized that a biostatistician should be included in this planning component unless individuals with strong statistical background were present. The plan should include consideration of quality assurance in all aspects of the study. A small scale pilot study was recommended if the research was unique and no reference material was available. Communication between all research participants was emphasized as a critical requirement to ensure success.

The documentation of procedures was also well recognized as a critical requirement for effective quality assurance. Procedures carried out in all aspects of field and lab sampling and analysis must be clearly detailed. Training manuals are desirable and should be clear and precise so any new individual approaching the research can easily identify techniques. It was also recognized that training programs are required for technical aspects of the program including field and laboratory procedures. A mechanism to test effectiveness of training techniques is also recognized as an important aspect to schedule into the research design.

To complete the QA plan an effective reporting format is required. The intent is to ensure that review occurs throughout the research schedule. Mechanisms to ensure specifications for data requirements are being met and that progress is being tracked are necessary to be confident in the final data. Such mechanisms are also critical to flag errors and deficiencies early in the research schedule. In addition it is recommended that a mechanism to provide feedback from the client be established. A strategy is needed to ensure progress is being made toward the principal objectives agreed upon with the client.

As a result of working through the quality assurance requirements in the case study exercise, the working group identified two critical issues in assessing terrestrial research. The first is how to ensure quality is integrated into the planning phase of research and carried through all components of the study program. It was recognized that QA is well accepted by the scientific community in certain areas of science, in particular laboratory procedures. The scientific community, in the interests of good science, must increase its demand for quality assurance in all components of research before accepting scientific data. The question of what QA criteria were used needs to be asked for each facet of a study. Ultimately QA knowledge should be reviewed before approving submissions to the scientific literature or acceptance in scientific reports.

The academic community must also carry the QA torch into its scientific endeavours. This vehicle is critical to ensure that education and training in scientific principles is married with quality assurance to ensure precision and accuracy in each phase of research. The scientific community should be striving with the academic community to promote and foster sound quality assurance practices throughout all scientific disciplines.

The third community involved in the promotion of integrated quality assurance criteria with planning protocols is that of management. The promotion of quality assurance is essential to ensure a high degree of meaningful performance and results from employees. It was recognized that management in its pursuit of a quality product and in the protection of its reputation for such, should be on the QA bandwagon. However, it was also recognized that the implementation of quality assurance programs is expensive in both time and resources. For this reason it may not always receive full management support.

To ensure quality assurance is well accepted by management, it was suggested that it become more closely associated with performance measurement of both programs and associated staff. To this end managers may incorporate QA criteria in performance standards or guidelines. These criteria could be presented up-front during recruitment and serve as performance indicators in regular appraisals throughout employment. Alternately, the suggestion was made that for scientists or scientific support staff quality assurance criteria should be actively promoted as a desired performance tool. In other words, as individuals the scientists should be promoting quality assurance integration in performance assessment.

The second critical issue for quality assurance in terrestrial ecosystems is the integration of QA data with results data in reporting. At present, quality assurance information is not routinely presented or expected. This is no longer appropriate and must be revised to incorporate both experimental results and QA data as the acceptable procedure in science. Questions of data confidence, precision and accuracy must be rigourously pursued before accepting scientific results.

Although consensus was not obtained in selecting the most desirable approach to reporting quality assurance data and scientific results, several approaches were suggested. These included an audit system for quality assurance and quality control data which could be appended to data sets. Alternately, a selected series of key quality assurance indicators could be presented in the reporting format with more detailed information available in an appendix or by contacting the principal investigators. These key indicators of accuracy and precision for example could appear as automatic headers or footers standardized to graphs or tables. Detailed information on all quality assurance criteria could be routinely stored with the data base and accessible to interested parties on request.

Summary

In summary the critical issues identified by the terrestrial workgroup were as follows:

1. How to ensure quality assurance is integrated into planning and development of scientific research

   . scientific community
   . academic community
   . management and supervision

2. How to append quality assurance data to results

   . audit quality assurance and quality control
   . select quality assurance indicators
   . automatic headers with key QA data
   . detailed QA on database storage files

Case Study - Quality Assurance in Terrestral Ecosystems

I Field Component

Define Objective:

- Soil solutions collected from Sugar Maple and Red Spruce Sites in an aera impacted by high acidic deposition (740 kg/ha/yr SO_4) through lysimeters

- Soil solutions chemically analyzed to determine chemical characterization and to develop fertilization criteria for a controlled experiment using spruce and maple seedlings grown in closed columns

Determine QA Activities:

Critical Questions from Error Analysis

1. Field Collection could result in contamination by:

   - tube contamination
   - plate contamination
   - transport method
   - lysimeter not stabilized
   - handling by field staff

2. Field Collection by Lysimeter requires protocol such as:

- check lysimeter is clean in lab
- develop stabilization procedure
- blanks for transport and collection check
- check out characteristics of lysimeter plates
- stabilize desired parameters on lysimeter
- ensure sufficient sample volumes are calculated

3. Sample Bottles

- ensure clean prior to taking to field
- develop labelling procedure and follow-up
- incorporate blank samples

4. Collection Procedure

- document procedures in manual
- provide training for field staff
- ensure follow-up, possibly testing program
- document logging procedures and follow-up

5. Holding time

- ensure a check against desired parameters
- develop clear guidelines, i.e. Chain of Custody

6. Transportation

- ensure labelling, storage and handling procedures are documented

7. Subordinate Information

- collect any relevant data as to weather conditions, time of day, or any other useful information

II Experimental Component

Define Objective:

- To quantitatively evaluate fertilization as a component of decline in sugar maple and red spruce seedlings grown in controlled conditions.

- To use soil columns treated with solutions typical of natural soil solutions from an area of high acidic deposition to evaluate growth and development of sugar maple and red spruce seedlings treated with Acidic fog, Varied CO_2 content, Varied aluminum content and Varied fertility levels.

Determine QA Activities

Critical Questions from Error Analysis:

1. Controlled Environment

   - Ensure medium appropriate
   - Ensure random distribution
   - Ensure sand doesn't compact in columns
   - Equilibration of sand in tubes is a slow process

2. Seedlings Selection

   - Do seedlings selected provide a suitable vector to determine a nutrient deficiency effect?
   - Is Variability in seed source accounted for?
   - Is Variability in rearing seed accounted for?
   - Is Variability in experimental medium accounted for?

3. Column and Sand

   - Placement of sand in columns needs to be standardized
   - Chemical characteristics of sand needs to be calibrated
   - pH in sand columns requires equilibrating

4. Growth Rate of Trees

   - Is generic variability accounted for?
   - Standardize criteria for physical measurements
   - Document sampling procedures in a manual

5. Discolouration

   - This subjective measure is difficult to standardize
   - May use colour charting, low accuracy & precision
   - Best to establish broad, distinct criteria
   - Training for repeatability required

6. Fine Root Turn-Over

- Account for genetic variability
- Culture variance may occur
- Ensure adequate replication
- Measure overlap and agree on estimate

7. Photosynthesis Measurement

- Instrument measurement
- Ensure calibration process well documented
- Provide training on use and calibration

8. CO_2 Treatment

- Calibration of CO_2 analyzer routine
- Check distribution of gas inside chamber
- Incorporate an audit and control chart system
- Plot target concentrations against actual achieved

9. Acid Fog Treatment

- Calibration of pH meter critical
- Control chart pH readings prior to applications
- Tracking system must be incorporated for acid reservoirs, delivery system and distribution of acid.

WORKGROUP 5: QUALITY MANAGEMENT FOR TOXIC ORGANICS IN AIR AND PRECIPITATION

Participant List

Neville Reid (Leader)
Air Resources Br
Ontario Ministry of the Environment,
880 Bay St., 4th Floor,
Toronto, Ontario, M5S 1Z8
Phone (416) 326-1691
FAX (416) 326-1773

Bill Strachan (Rapporteur)
Lakes Research Branch
Natural Water Research Institute
Environment Canada
CCIW, 867 Lakeshore Rd.,
Burlington, Ontario, L7R 4A6
Phone (416) 336-4775
FAX (416) 336-4989

Haig Agemian
National Water Quality Laboratory,
Environment Canada,
CCIW, 867 Lakeshore Rd.,
Burlington, Ontario, L7R 4A6
Phone (416) 336-4679
FAX (416) 336-6404

Nabil Arafat
Research and Applications Branch
National Water Research Institute
Environment Canada
867 Lakeshore Rd.,
Burlington, Ontario, L7R 4A6
Phone (416) 336-????
FAX (416) 336-4989

Keijo Aspila
Research and Applications Branch
National Water Research Institute
Environment Canada
867 Lakeshore Rd.,
Burlington, Ontario, L7R 4A6
Phone (416) 336-4638
FAX (416) 336-4989

Jack Balsinger
Midwest Research Institute,
425 Volker Blvd.,
Kansas City, MO 64110
Phone (816) 753-7600
FAX (816) 753-8420

James Doull
Water Quality Branch,
Environment Canada
P. O. Box 861,
Moncton, N.B., E1C 8N6
Phone (506) 857-6606
FAX (506) 857-6608

Brian Foster
Laboratory Services Br.,
Ontario Ministry of the Environment,
125 Resources Rd.,
P.O.Box 213,
Rexdale, Ontario, M9W 5L1
Phone (416) 235-6002
FAX (416) 235-5744

Donald E. King
Laboratory Services Br.,
Ontario Ministry of the Environment,
125 Resources Rd.,
P.O.Box 213,
Rexdale, Ontario, M9W 5L1
Phone (416) 235-5838
FAX (416) 235-5744

John Lawrence
Research Application Branch,
National Water Research Institute
Environment Canada
867 Lakeshore Rd.,
Burlington, Ontario, L7R 4A6
Phone (416) 336-4927
FAX (416) 336-4989

Richard Turle
Canadian Wildlife Service,
Environment Canada,
100 Gamelin Blvd.,
Hull, P.Q., K1A 0H3
Phone (613) 997-1412
FAX (613) 953-6612

Helmut Winter
Environment Quebec,
Complexe scientifique,
2700 rue Einstein,
Sainte-Foy, P.Q., G1P 3W8
Phone (418) 643-8225
FAX (418) 643-9023

QUALITY MANAGEMENT FOR TOXIC ORGANICS IN AIR AND PRECIPITATION

Workgroup Objectives

To develop a QA/QC issue related statement on toxic organic compounds in the air and in precipitation as sampled in a regional rather than local scale network (i.e., hundreds of kilometers in grid intervals).

QA/QC Approaches

A number of quality management issues should be addressed when establishing and operating a network to describe the regional distribution of toxic organic chemicals in atmospheric samples. The resolution of these issues should lead to the preparation of a Feasibility Report, a Project Plan and a Quality Management Plan. Such documents would then provide the basis for program approval and successful implementation.

The network objectives of the program must be defined but they are here presumed to be a description of levels of selected chemicals in atmospheric matrices as required to provide loading estimates to the terrestrial and aquatic compartments of the environment. To achieve this, it will be necessary to establish the scale of the network, i.e. regional with a smaller scale if necessary. A feasibility study(ies) based on existing literature and/or field work would be required to answer questions such as:

- spatial and temporal variation to be expected in levels of analytes;
- capability of the sampling methodology (reliability, suitability and performance specifications);
- comparability of sampling methodologies (co-locating samplers, spiking and recovery experiments);
- sensitivity of analytical systems (will affect sample size and frequencies);
- spatial resolution of site locations (affected by all sources of uncertainty above);
- criteria for local selection of sites;
- target chemical species to be investigated;
- identification of data quality objectives;
- resource requirements including those for the quality management.

Given this information, a network can be designed to satisfy the objectives identified. It should be stressed that for an atmospheric network, there will be a need for equipment at each site to provide local meteorological information. This is required to permit interpretation of the results and should include, at least temperature, relative humidity, wind speed and direction and precipitation. Longer air movements may also be required from a continental scale meteorological program.

Frequent review and evaluation of the data against predefined acceptability criteria, including its quality assurance, is required frequently in order to implement changes in the program necessary to achieve the stated program objectives. Once a network is established, a number of considerations are necessary to ensure its successful operation if it is to produce data of the needed quality. Some of these are discussed below.

Field Site Activity

A program addressing quality assurance in the field should be incorporated in the operating program of the network. It should include documentation of procedures and a training program for all site operators as well as for those involved with the transport of the collected samples. There should be regular checks on the sample collection procedures by employing blanks to determine the impact of handling and transport. In order to provide for comparability between different sites and/or networks, there should be replicate sampling including co-location of samplers; performance audit samples should be used where possible. Schedules for equipment maintenance and operational audits should be established; there should be regular checks to determine whether specifications for equipment and operator performance are being met. As part of the field documentation, there should be an identified remedial action plan and the highest priority assigned to correcting any deficiencies. There should be periodic site checks to determine whether the particular location continues to be suitable.

Sample Storage/Transport

Quality control checks should be undertaken on a regular basis to determine the sample integrity during transport and storage. To achieve this, sample blanks and spiked samples are required. Quality assurance aspects relevant to storage and transport may include: refrigeration, the use of preservatives, dark storage, protection from physical damage and chemical contamination, determination of maximum storage times for samples.

Chemical Analysis

Quality control procedures for analytical laboratories are moderately well described and, on the whole, individual laboratories dealing in routine monitoring of environmental samples have well-established protocols. Such controls will include the use of reference materials and calibration standards as well as frequent processing of

blanks, blind reference samples, replicate analyses and analysis of spiked samples. The laboratories involved in the atmospheric network will have regular maintenance schedules for their equipment together with a budget and plan for rectifying any malfunctions. One approach to ensure the condition of the analytical system is the use of control charts. Particularly where there are several laboratories involved in the network, round-robin exercises to establish comparability of data will be an important part of the analytical QA program. All methods should be documented and any changes recorded and comparisons made between results with procedures both "before" and "after" such changes. As with all of the separate aspects identified here for QA/QC in atmospheric monitoring, regular audits to ensure that the established procedures are being followed should be made.

Data Management

Although the preceeding is intended to generate quality data, there also remains the problem of record retention and use for the established purposes and others which may arise. All information (eg. results, QA/QC data, non-quantitive information from sites, etc.) relevant to the program should be retained and should be available in a unified data base. Verification of data entry is critical; it is the responsibility of the recorder to check back with the originator to ensure the data is correctly entered and stored. This will include checks on sample identifiers/coding and a regular check for correspondence among primary (i.e., field), laboratory and recorded data. Attention should be paid to correct units and decimal transposition. Information in a graphical form should be prepared and examined for anomalies. Quality assurance should include validation of the data transfer and any data calculations. Pattern recognition (established proportions of two or more of the analytes) provide one means of flagging potentially false data although "outliers" should not be rejected per se. Timely review of the data in this and other ways should be undertaken in order to initiate remedial actions before large numbers of samples are destroyed or unnecessary data generated.

Summary

Data quality management for atmospheric sampling networks should include detailed documentation of procedures and all observations including those from the field, during transport and in the laboratory. Training of field personnel was also considered highly important. Auditing of conformation to established procedures and timely inspection of data should provide feedback to the operation of the network including all aspects of sample collection and analysis and data storage. A variety of measures to achieve this are presented.

WORKGROUP 6: QUALITY MANAGEMENT FOR MONITORING TRACE TOXIC ORGANIC CONTAMINANTS IN AN AQUATIC ECOSYSTEM

Participant List

Carol-Ann Manen (Rapporteur)
NOAA National Ocean Service
6001 Executive Blvd., Room 312
Rockville, MD 20852
Phone (301) 443-8655
FAX (301) 231-5764

Dave Sergeant (Leader)
Fisheries and Oceans Canada
867 Lakeshore Road, P.O. 5050
Burlington, Ontario, L7R 4A6
Phone (416) 336-4863
FAX (416) 336-4819

Michael Comba
Lakes Research Branch
National Water Research Institute
Environment Canada
867 Lakeshore Road, P.O. 5050
Burlington, Ontario, L7R 4A6
Phone (416) 336-4617
FAX (416) 336-4989

Valerie Ervin
HJF Forestry Centre
Forestry Canada
P.O. Box 4000
Fredericton, NB, E3B 5P7
Phone (506) 452-3562
FAX (506) 452-3525

Dominique Duval
St. Lawrence Centre Laboratory
Environment Canada
1001 Pierre-Dupuy Street
Congueuil PQ, J7K 1A1
Phone (514) 646-6042
FAX (514) 646-1728

Robert Graves
U.S. Environmental Protection Agency
Environmental Monitoring Systems Laboratory
26 W. Martin Luther King Drive
Cincinnati, OH 45219
Phone (513) 569-7325
FAX (513) 569-7424

Hari S. Samant
Environmental Protection, Atlantic Region
Environment Canada
45 Alderney Drive, 5th Floor Queen's Square
Dartmouth, NS B2Y 2N6
Phone (902) 426-6237
FAX (902) 426-2690

Tom Pollock
Water Quality Branch
Environment Canada
Federal Bldg., Main Street, P.O. 861
Moncton, NB, E1C 8N6
Phone (506) 857-6606
FAX (506) 857-6608

Michael Forbes
Environment Canada
National Water Quality Laboratory
867 Lakeshore Road, P.O. 5050
Burlington, Ontario, L7R 4A6
Phone (416) 336-4577
FAX (416) 336-6404

Deborah Coffey
NSI Technology Services Corporation
1600 SW Western Blvd.
Corvalis, OR 97333
Phone (503) 757-4666
FAX (503) 757-4335

Alfred S.Y. Chau
National Water Research Institute
Environment Canada
867 Lakeshore Road, P.O. 5050
Burlington, Ontario, L7R 4A6
Phone (416) 336-4653
FAX (416) 336-4989

Ghulam H. Jamro
National Water Quality Laboratory
Environment Canada
867 Lakeshore Road, P.O. 5050
Burlington, Ontario, L7R 4A6
Phone (416) 336-4639
FAX (416) 336-4989

QUALITY MANAGEMENT FOR MONITORING TRACE TOXIC ORGANIC CONTAMINANTS IN AN AQUATIC ECOSYSTEM

Workgroup Objective

The group considered a hypothetical but typical situation as a rationale for monitoring, briefly addressed the quality management issues that would be expected to arise, and suggested approaches to these issues.

Scenario

An environmental group "Blackedge" notices a somewhat disreputable dredging company "Seldom Seen Dredging Co" dumping dredge spoils into the eastern end of Lake Erie. They follow the operation and notice that spoil is coming from Fighting Island and next notice that fish at the spoil dump site are dying. Analyses of 10 of the dead fish indicate that 5 of them contain 20 times the Canadian Health and Welfare (HW) and U.S. FDA action level of 10 ppt in edible fish flesh for compound Q, an acutely toxic, non-volatile synthetic organic compound.

Blackedge immediately goes to the media, T.V., news, etc. and the federal and local governments bordering Lake Erie about the dangers of the dredge spoils. The various Canadian and U.S. federal and local governments appoint Workgroup 6 to determine if this is a threat to:

- Human Health
- Fishery Resources
- Environment

Problem(s) (International in scope)

1) Are the levels in any fish (perch, trout, walleye, smelt) sufficiently high in any species to close the fishery for that species?

2) Are the levels in all 4 species sufficiently high (over 10 ppt) in all 4 species to close the lake to commercial and sport fishing?

3) Are the levels sufficiently high and widespread to indicate a need for damage assessment?

Approaches

- To determine if the levels of Q in any or all the edible portion of fish flesh of perch, trout, walleye, smelt are > 10 ppt (mean, max, percentiles 1/10, 1/100).

- To collect the above data in such a fashion as to be legally defensible.

o Design to achieve acceptable resolution in comparing exposed and control sites, and comparison to residue criteria.

Initial consideration was given to compositing extensively in order to reduce analytical costs, however, as part of the DQO process, it was recognized that variability among composites, due to variable sex and age composition, would likely be too great to achieve the desired resolution.

Methods: Design

- Historical data on variability of fish populations and analyses for Q in fish flesh provided to statistician to determine number of samples for adequate resolution.

- Two sampling locations; exposed and control; match by sex, age, etc. in paired design.

- Composite similar individuals as required to obtain adequate sample volume in small size classes; otherwise analyse individual fish.

Method: Collection (Canada)

- Teams experienced in the collection of fish for organic contaminant analysis

- Standard Operating Procedures (SOPs)

    - collection procedures, equipment, documentation includes net size, trawl duration, location, number of species

    - sorting, taxonomy, weight, length, sex, gross pathology, shipping, holding, freezing

    - chain of custody

- Checklist for field operations

- Briefing of field crew

    - objectives of sampling
    - hazards of Q (MSDs)

Methods: Sample Regulation SOPs (Canada)

- Maintain sample integrity
- Maintain chain of custody
- Initiate sample tracking

Methods: Preparation (Canada)

- Species specific, edible portions, preparation methods
- Homogenize and aliquot to obtain duplicates

Methods: Analyses (International)

- Unspecified methods
- Supply calibration standards, SRMs and samples
- Supply blind duplicates
- Specify analytical performance bounds (recovery, detection, accuracy, precision)
- Specify reporting format

Reporting

- Corrected or uncorrected data, identified as such, with associated blanks, recoveries and lipid contents.
- Analytical method: extraction and analysis, including operational parameters (data acceptability criteria) as per SOPs.
- Detection Limits stated and defined
- Data reduction and verification methods, software etc.
- Raw Data: chromatograms of SRM and sample (for a species showing Q)
- Comments and observations

Data Interpretation (Workgroup 6)

- Examination of QA data to determine if results meet DQOs (within the bounds of variability used in design).

- Unsuitable data dropped or sample reanalyzed.

- Statistical treatment to check for fish age/size effects within species.

- Statistical treatment to determine if data sets can be combined (across species) for site/control comparison. Outlying data dropped.

- Statistical comparison of combined data or species data as appropriate (across sites); e.g. comparison of size vs. residue relationships across sites.

Summary

Typically residue comparisons between exposed and reference sites, and of site descriptors (means, percentiles) to regulatory criteria are required in fisheries management for human health protection. In making such comparisons it is critical to design for acceptable resolution, and to consider age, sex and species effects at the

design stage and in data interpretation. Data quality criteria should be defined at the outset and reported with the results and associated data quality measures. The DQO process was effective in permitting recognition of design problems prior to study implementation, and correction of the design to achieve study objectives.

APPENDIX A

Questionnaire and Summary

Third Ecological Quality Assurance Workshop
Burlington, Ontario

April 24 -26 1990

EVALUATION SHEET

Summary of information culled from evaluation sheets
submitted by 41 of the 75 participants
[25 from USA, 2 from Europe, 48 from Canada]

1. Indicate the workgroup in which you participated.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ none

** Evaluation sheets were submitted by 41 participants distributed in workgroups as follows: **

6 9 8 7 5 6

2. How applicable was the subject matter in this workshop to your job?

Wrkgrp #	☐ significant	☐ adequate	☐ insignificant
1	5		
2	7		
3	6	2	
4	5	2	
5	3	2	
6	3	2	
		3	
	29	11	

3. **Views on the format of the workshop:**

i. **time allocated for discussions in plenary session;**

Wrkgrp #	☐ adequate	☐ too long	☐ too short
1	2	3	1
2	4	1	3
3	2	2	4
4	4		3
5	2		3
6	5		1
	19	6	15

ii. **time allocated for workgroup discussions;**

Wrkgrp #	☐ adequate	☐ too long	☐ too short
1	6		
2	8		1
3	6		2
4	5		1
5	2	1	2
6	2	1	3
	29	2	9

iii. **how effective was the workgroup format?**

Wrkgrp #	☐ excellent	☐ good	☐ poor
1	6		
2	2	7	
3	1	6	1
4	5	2	
5		5	
6		5	1
	14	25	2

iv. the relevance of the topics covered in presentations by invited speakers to the topics discussed in workgroup sessions;

Wrkgrp #	☐ excellent	☐ good	☐ poor
1	2	4	
2	1	8	
3		4	3
4	2	5	
5	1	4	
6		5	
	6	31	3

v. suggestions for improving the format of similar workshops.

- More time needs to be allocated to discussions, both in plenary and workgroup sessions; adding an extra day should be considered
  Additional time allocated to workgroup discussions would allow a second look at the output and a review of the draft report
  The output of workgroup discussions should be prepared in a short preliminary report and distributed to workgroup participants to allow more time for discussion within the workgroup on selected topics

- The presence of one invited speaker per workgroup would be desirable to bring his/her expertise into the discussions

- The use of QA case studies, particularly those that have gone through planning and/or audit stages

- Statistical aspects of ecological QA and accuracy of biological measurements need to be addressed

- The workshop should be focused on different aspects of QA/QC logic and uses, e.g., in areas such as project design criteria, legal/remedial actions, who should decide on QA/QC parameters

- Aggressive and experienced facilitators should be used for leading workgroup discussions

- Improvement in the output of workgroups would be gained by using effective discussion leaders and preparing good outlines of principal issues to be discussed

- More time needs to be allocated to the wrap-up session on workgroup discussions , possibly by shortening the 1st day plenary session and/or the length/number of invited presentations
  Better control of time allocated to invited speakers is needed to ensure that adequate time is available for questions and answers
  The time allocated to each invited should be adjusted in function of the material to be presented
  Invited speakers' presentations should be more focused on QA aspects in their respective fields, with appropriate emphasis on problems encountered, lessons learned, etc.
  Invited presentations should go beyond "preaching to the converted"

- The entire workshop should be focused on one theme, with participants invited to contribute their special expertise to address particular aspects during discussions

- Material on the topics to be discussed during the workshop should be prepared and made available prior to the workshop

- The objectives of the workgroup discussions need to be well defined at the very beginning of the discussion

- More field personnel and managers should be invited to participate to the workshop so to ensure better dissemination of the "good word" back home; adding sessions in the field would be beneficial and should be considered

- A social hour/mixer should have been planned in the early part of the workshop to allov participants to meet each other outside of the workshop sessions

4. i. What were your objectives for attending the workshop?

   ii. Were your objectives met?

Most of the participants that stated objectives for attending the workshop indicated that these were met. Typical objectives were:
- interactions with individuals involved in different aspects of QA
- update on ideas/approaches to QA
- learn on specific aspects such as QA/QC for bioassays, application of QA to environmental assessment, QA/QC elements of environmental sampling, etc.

5. Comments on strong points of workshop.

- The diversity of experience/expertise found in the workgroups [this comment vas the one offered the most often]

- The workshop vas well organized and located in excellent facilities

- Good teams of leaders-rapporteurs vere assembled for this workshop [several participants made this comment]

- The workshop provided a good opportunity to establish contacts with individuals involved in QA matters

- The noticeable improvement in the running of the workgroups could be due to the participants and/or the unstructured format of the discussion groups

- The size of the workgroups vas optimal

- The workshop provided a good overviev of important aspects of ecological QA and the addition of an international perspective vas a definite plus

- The workshop constitutes an important contribution for exchanges and progress tovard a standardization of approaches to data quality management

- The improved coverage of biological aspects of ecological QA over previous years vas appreciated

- Some of the invited speakers were good and effective in putting across good ideas on QA
  Non-traditional points about QA were presented and emphasis was placed on the need to involve management in the early phases of QA/QC planning

6. **Comments on weak points of workshop.**

- There was too much overlap in the concepts covered by the invited speakers in their presentations

- There was a lack of focus in the presentations made by invited speakers
  More focus needed in invited presentations to address the implementation and the results of QA/QC programs
  There were irrelevant presentations made, with more questions raised than answers offered

- The audience for these workshops will have to be changed to include working scientists and managers in support of integration efforts

- Too much time was spent on generalizations, skipping over specific issues; workgroups need a more structured format and firm organization

- The outlines covering the topics to be discussed in workgroups did not contain sufficiently detailed guidelines

- There was a lack of linkages between the invited presentations and the topics to be discussed in workgroup sessions

- The workshop did not provide a resolution of specific issues currently affecting results or their determination;
  The workshop did not result in the preparation of a summary or of common recommendations to be reflected in the Proceedings

- The lack of a distillation or overall synthesis of the output of the workgroup discussions made available in a preliminary form during the workshop means that participants have to wait for the availability of the Proceedings document to find out, hopefully, what were the points of agreement, the points of disagreements and the issues in need of further discussion

- The workgroup reports were too general and therefore contained excessive repetitions of the same points

- Too much time was spent for finalizing and revising the rapporteur's final report

- There is a need to diffuse the output of the workshop to audiences beyond workshop participants

- There is a need to define the goals of future workshops with a long-term vision

- The 1st day plenary session was too long [several participants made this comment]

- The absence of representatives from the private industry where QA is practised constitutes a weak point

7. **If a similar workshop were held next year, would you recommend this workshop to your colleagues?**

☐ highly recommend ☐ recommend ☐ not recommend

Wrkgrp #		
1	4	2
2	5	4
3	1	7
4	4	3
5	4	1
6	5	1
	23	18

8. **What focus would you like to see for the next Ecological QA Workshop?**

- Specific case studies should be included as topics for discussion
  Examples of actual programs that have implemented good QA procedures should be presented for study and critique
- Statistical aspects of QA need to be addressed
- Workgroups could be given the task of designing a monitoring program, with the participants contributing examples of QA/QC aspects drawn from their work experience
- The focus on biological aspects of QA should be continued, for instance the relevance of chemical data to the evaluation of biological communities and the QA issues related to this type of evaluation
- The topics discussed in workgroup no. 1 should be discussed further
- The use/usefulness of QA/QC; the specifics of QA/QC measures; data interpretation with QA/QC influence
- Taxonomic QA; the design of standard operating procedures
- Quality Control versus Quality Assessment; more focus on research-type ecology; what needs to be reported in terms of QA/QC
- Development of QA sample types and protocols; extend D. Hart's check list concept
- Estimating sources of error or variance in a programme and design of QA/QC measures to reduce or contain these uncertainties
- Integration of QA into good management and hov to overcome the natural barriers at the program/research management level
  Continued integration of QA into data reporting/project design and the writing of proposals
- Preservation and homogenization of samples; preparation, homogeneity and stability of "in-house" reference materials for organics or major ions as part of analytical programs focused on industrial effluents, sevage, sediments and biota

- Definition and determination of errors at each stage of a monitoring or survey program
- Management needs to be brought on stream and emphasis placed on the integration of procedures and methods and the linkage between QA and results under a data management approach
- Training and organization aspects of QA need to be addressed
- Methodologies and their comparability needs to be discussed, with particular attention to the problems encountered in air monitoring programs
- QA and specimen banking
- Fine tuning of the DQO process should be continued as it pertains to decision making regarding the measurements of organic/inorganic parameters at different levels
- QA/QC requirements in support of status and trends evaluation programs Long-term monitoring programs and their QA considerations: how to ensure that changes in the course of the program will not adversely affect the outcome, i.e., the detection of long-term trends

9. **Indicate your preference for the location of the next Ecological QA Workshop.**

U.S.A.

Wrkgrp #	☐ East Coast	☐ Central	☐ West Coast
1	1	1	
2	1	2	1
3	1	1	2
4	1		1
5	1		
6		2	
	5	6	4

Canada

Wrkgrp #	☐ East Coast	☐ Central	☐ West Coast
1	1	1	1
2		1	2
3			1
4	1		
5			2
6			2
	2	2	8

APPENDIX B

Quality Assurance Glossary

DRAFT

QUALITY ASSURANCE GLOSSARY AND ACRONYMS

Quality Assurance Management Staff (RD-680)
Office of Modeling, Monitoring Systems and Quality Assurance
Office of Research and Development
United States Environmental Protection Agency
401 M Street, SW
Washington, DC 20460

July 23, 1990

DRAFT

GLOSSARY OF TERMS

Absolute method - a body of procedures and techniques for which measurement is based entirely on physically defined, fundamental quantities.

Acceptable quality level - a limit above which quality is considered satisfactory and below which it is not. In sampling inspection, the maximum percentage of defects or failures that can be considered satisfactory as an average.

Acceptable quality range - the interval, between specified upper and lower limits of a sequence of values, within which the values are considered to be satisfactory.

Acceptable value - an observed or corrected value that falls within the acceptable range. See Corrected value and Observed value.

Acceptance sampling - the procedure of drawing samples from a lot or population to determine whether to accept or reject a sampled lot or population.

Accepted reference value - a numerical quantity that serves as an agreed-upon basis for comparison, and which is derived as; 1) a theoretical or established quantity based on scientific principles, 2) an assigned value, based on experimental work of some recognized organization, or 3) a consensus quantity based on collaborative experimental work under the auspices of a scientific or engineering group.

Accreditation - a formal recognition that a laboratory is competent to carry out specific tests or specific types of tests. See Certification.

Accreditation criterion - a requirement that a laboratory must meet to receive authorization and approval to perform a specified task.

Accredited laboratory - a laboratory which has been evaluated and given approval to perform a specified measurement or task, usually for a specific parameter and a specified period of time.

Accuracy - the degree of agreement between an observed value and an accepted reference value. Accuracy includes a combination of random and systematic error or bias components which are due to sampling and analytical operations; a data quality indicator.

Aliquant - a subsample derived by a divisor that divides a sample into a number of equal parts but leaves a remainder; a subsample resulting from such a divisor. See Subsample.

Aliquot - a subsample derived by a divisor that divides a sample into a number of equal parts and leaves no remainder; a subsample resulting from such a division. In analytical chemistry the term aliquot is generally used to define any representative portion of the sample.

Alpha error - see "Type I Error."

DRAFT

Alternate method - any body of procedures and techniques of sample collection and/or analyses for a characteristic of interest which is not a reference or approved equivalent method but which has been demonstrated in specific cases to produce results comparable to those obtained from a reference method.

Analysis - the determination of the qualitative and/or quantitative composition of a substance.

Analyte - the substance, a property of which is to be measured, in a chemical analysis.

Analytical batch - a group of samples, including quality control samples, which are processed together with the same method, the same lots of reagents, and at the same time or in continuous, sequential time periods. Samples in each batch should be of similar composition and share common internal quality control standards.

Analytical blank - see Laboratory blank.

Analytical limit of discrimination - see Method detection limit.

Analytical reagent (AR) - the American Chemical Society's designation for the highest purity of certain chemical reagents and solvents. See Reagent grade.

Arithmetic mean - the sum of all the values of a set of measurements divided by the number of values in the set, usually denoted by $\bar{x}$; a measure of central tendency. See Measure of central tendency.

Assignable cause - a factor or an experimental variable shown to significantly change the quality of an effect or a result.

Audit - a systematic evaluation to determine the quality of some managerial or operational function or activity. See Audit of data quality, Management systems audit, Management systems review, Performance evaluation audit, and Technical systems audit.

Audit of data quality (ADQ) - a qualitative and quantitative evaluation of the documentation and procedures associated with environmental measurements to verify that the resulting data are of acceptable quality.

Audit sample - a blind sample used for an audit. See Blind sample.

Average - see Arithmetic mean.

Background level (non-instrumental) - the concentration of substance in a defined area during a fixed period of time before or after a data gathering operation.

Batch - a quantity of material produced in one operation, considered to be a uniform discrete unit.

DRAFT

Batch-lot - the samples collected under sufficiently uniform conditions to be processed as a group. See Batch, Batch size.

Batch-sample - one of the samples drawn from a batch.

Batch-size - the number of samples in a batch-lot.

Beta error - see Type II Error.

Bias - systematic error manifested as a consistent positive or negative deviation from the known or true value. It differs from random error which shows no such deviation.

Blank sample - a clean sample or a sample of matrix processed so as to measure artifacts in the measurement (sampling and analysis) process.

Blind sample - a subsample submitted for analysis with a composition and identity known to the submitter but unknown to the analyst and used to test the analyst's or laboratory's proficiency in the execution of the measurement process. See Double-blind sample.

Bulk sample - a sample taken from a larger quantity (lot) for analysis or recording purposes.

Calibrant - see Calibration standard.

Calibrate - to determine, by measurement or comparison with a standard, the correct value of each scale reading on a meter or other device, or the correct value for each setting of a control knob. The levels of the calibration standards should bracket the range of planned measurements.

Calibration-check - see Calibrate.

Calibration-check standard - see Calibration standard.

Calibration curve - the graphical relationship between the known values for a series of calibration standards and instrument responses.

Calibration drift - the difference between the instrument response and a reference value after a period of operation without recalibration.

Calibration standard - a substance or reference material used to calibrate and instrument.

Candidate method - a body of procedures and techniques of sample collection and/or analysis that is submitted for approval as a reference method, an equivalent method, or an alternative method.

Carrying-agent - any diluent or matrix used to entrain, dilute or to act as a vehicle for a compound of interest.

DRAFT

CAS# - Chemical Abstracts Service registry number of elements, chemical compounds, and certain mixtures.

Cause-effect diagram - a graphical representation of an effect and possible causes. A popular one is the Ishikawa "fish bone diagram."

Central line - the line representing the arithmetic mean or an expected value of the standard measure that is plotted on a control chart. See Control chart.

Certification - the process of testing and evaluation against specifications designed to document, verify, and recognize the competence of a person, organization, or other entity to perform a function or service usually for a specific time. See Accreditation.

Certified Reference Material (CRM) - a reference material having one or more of its property values established by a technically valid procedure and accompanied by or traceable to a certificate or other documentation issued by a certifying body. See Certification and Reference material.

Certified value - the reported numerical quantity that appears on a certificate for a property of a reference material.

Chain-of-custody - an unbroken trail of accountability that insures the integrity of samples, data and records.

Chance cause - an unpredictable, random determinant of variation of a response in a sampling or measurement operation.

Characteristic - see Property.

Check sample - an uncontaminated sample of a matrix spiked with analyte(s) from an independent source in order to measure the concentration of analyte(s) in a sample or the performance of an analytical method.

Check standard - see Calibration check standard.

Chi-square test - a statistical test of the agreement between the observed frequency of binomially distributed events and the frequency expected according to some hypothesis.

Clean sample - a sample of a natural or synthetic matrix containing no detectable amount of the analyte of interest and no interfering material.

Coefficient of variation (CV) - a measure of relative dispersion (precision.) It is equal to the standard deviation divided by the arithmetic mean and multiplied by 100 and expressed as a percentage value. Also called relative standard deviation.

Collaborative testing - the evaluation of an analytical method by typical or representative laboratories using subsamples prepared from a homogeneous standard sample.

DRAFT

Collocated sample - one of two or more independent samples collected so that each is equally representative for a given variable at a given space and time.

Collocated samplers - two or more identical instruments usually operated simultaneously, to supply a series of duplicate or replicate samples for estimating precision.

Comparability - a measure of the degree to which different methods, data sets and/or decisions can be represented as similar; a data quality indicator.

Completeness - the amount of valid data obtained compared to the planned amount, and usually expressed as a percentage; a data quality indicator.

Component of variance - a part of the total variance associated with a specified source of variation.

Composite sample - a sample prepared by physically combining two or more samples or substances and processed to ensure homogeneity. See Flow-proportioned sample and Time-proportioned sample.

Confidence interval - a range that has a designated probability (i.e., confidence level) of including a specific population parameter.

Confidence level - the probability, usually expressed as a percentage, that a confidence interval will include a specific population parameter; confidence levels usually range from 90 to 99 percent.

Control chart - a graph of some measurement plotted over time or sequence of sampling, together with control limit(s) and, usually, a central line. See Action limit, Central line, Control limit and Warning limit.

Control limit - a specified boundary on a control chart that, if exceeded, indicates a process is out of statistical control, and the process must be stopped, and corrective action taken before proceeding; the boundary specified is usually three standard deviations (i.e., the 99.72% confidence level.)

Control sample - see Quality control sample and Check sample.

Control standard - see Calibration-check standard.

Corrected value - the observed value after correction for values of a blank.

Correlation - a measure of association between two variables.
See Correlation coefficient.

Correlation coefficient - a number between -1 and 1 that indicates the degree of linearity between two sets of numbers. The closer to -1 or +1, the stronger the linear relationship between the sets (i.e., the better the correlation.) Values close to zero suggest no correlation between the sets and imply independence.

DRAFT

Critical-toxicity range - the interval between the highest concentration at which all test organisms survive and the lowest concentration at which all test organisms die within the test period.

Daily standard - see Calibration check standard.

Data - facts or figures from which conclusions can be inferred.

Data quality - the totality of features and characteristics of data that bears on their ability to satisfy a given purpose; the sum of the degrees of excellence for factors related to data.

Data quality indicators - quantitative statistics and qualitative descriptors that are used to interpret the degree of acceptability or utility of data to the user. The principal data quality indicators are precision, accuracy, comparability, completeness, and representativeness.

Data Quality Objective (DQO) - qualitative and quantitative statements of the level of uncertainty that a decision maker is willing to accept in decisions made with environmental data.

Data reduction - the process of transforming raw data by arithmetic or statistical calculations, standard curves, concentration factors, etc., and collation into a more useful form.

Data set - all the observed values for the samples in a test or study; a group of data collected under similar conditions and which, therefore, can be analyzed as a whole.

Data validation - see Validation.

Datum - the singular of data. See Data and Value.

Defensible - the ability to withstand any reasonable challenge related to the veracity or truthfullness of laboratory documents and derived data.

Degrees of freedom - the total number of items in a sample minus the number of independent relationships existing among them; the divisor used to calculate a variance term, in the simplest cases it is one less than the number of observations.

Dependent variable - a variable that is measured when a controlled experiment is conducted.

Detection limit - the lowest concentration or amount of the component of interest that can be determined by a single measurement at a stated confidence level (i.e., level of uncertainty.) See Method detection limit.

Determination - the process of measuring a property of interest. See Test determination.

Diluent - a substance added to another to reduce the concentration and resulting in a homogeneous end product without chemically altering the compound of interest.

Dilution factor - the numerical value obtained from dividing the new volume of a diluted substance by its original volume.

DRAFT

Discharge - any pollutant or combination of pollutants added to the environment from any point or area source.

Document control - a systematic procedure for indexing documents by number, date and revision number.

Double-blind sample - a sample submitted to evaluate performance with concentration and identity unknown to the analyst. See Blind sample.

Duplicate - an adjective describing the taking of a second sample or performance of a second measurement or determination. Often incorrectly used as a noun and substituted for "duplicate sample." Replicate is to be used if there are more than two items. See Replicate.

Duplicate analysis or measurement - the analysis or measurement of the variable of interest performed as identically as possible on two subsamples of a sample. Duplicate analysis is used to evaluate analytical or measurement variance.

Duplicate sample - one of two samples taken from the same population and carried through all steps of the sampling and analytical procedures in an identical manner. Duplicate samples are used to assess variance of the total method including sampling and analysis. See Split sample.

Dynamic blank - a sample-collection material or device (e.g., filter or reagent solution) that is not exposed to the material to be selectively captured but is transported and processed in the same manner as the sample. See Field blank, Instrumental blank and Sampling equipment blank.

Dynamic calibration - standardization of both the measurement and collection systems using a reference material similar to the unknown. For example, a series of air-mixture standards containing sulfur dioxide of known concentrations could be used to calibrate a sulfur dioxide bubbler system.

Environmental sample - a sample of any material that is collected from an environmental source.

Environmentally related measurement - any assessment of environmental concern generated through or for field, laboratory, or modelling processes; the value obtained from such an assessment.

Equivalent method - any method of sampling and/or analysis demonstrated to have a consistent and quantitatively known relationship to a reference method under specified conditions.

Error - any discrepancy between a computed, observed, or measured quantity and the true, specified, or theoretically correct value of that quantity.

Error (measurement) - the difference between an observed or corrected value of a variable and a specified, theoretically correct, or true value.

Error function - see Normal distribution.

Experimental variable - see Independent variable.

DRAFT

External quality control - the activities which are routinely initiated and performed by persons outside of normal operations to assess the capability and performance of a measurement process.

False negative error - see Type II Error.

False positive error - see Type I Error.

Field blank - a clean sample not exposed to the sample stream, but carried to the sampling site, exposed to sampling conditions (e.g., bottle caps removed, preservatives added) and returned to the laboratory as a sample. Field blanks are used to check background and sample-handling procedures. See Dynamic blank and Sampling equipment blank.

Field duplicates - see Duplicate sample.

Field reagent blank - see Field blank.

Field sample - see Sample.

Flow rate - the quantity-per-unit time of a substance passing a point, plane, or space; for example the volume or mass of gas or liquid emerging from an orifice, pump, or turbine or moving through a point in a conduit or channel.

Flow-proportioned sample - a composite sample or subsample collected from a fluid system at a rate that produces a constant ratio of sample accumulation to matrix flow rate.

Full-scale response - the maximum output of a measurement instrument in a given range as displayed on a meter or scale.

Functional analysis - a mathematical evaluation of each component of the measurement system (sampling and analysis) in order to quantitate the error for each component. A functional analysis is usually performed prior to a ruggedness test in order to determine those variables which should be studied experimentally.

Geometric mean - the antilogarithm of the mean of the logarithms of all the values in a set.

Good laboratory practices (GLP) - either general guidelines or formal regulations for performing basic laboratory operations or activities that are known or believed to influence the quality and integrity of the results.

Goodness-of-fit - the measure of agreement between the data in a data set and the expected or hypothesized values.

Grab sample - a single sample which is collected at one point in time and place.

Gross sample - see Bulk sample.

Homogeneity - the degree of uniformity of structure or composition.

DRAFT

In-control - a condition indicating that performance of the quality control system is within the specified control limits, i.e., that a stable system of chance is operating and resulting in statistical control. See Control chart.

Independent variable - a variable that is set at pre-selected level when a controlled experiment is conducted.

Inspection criterion - the specification(s) and rationale for rejecting and accepting samples in a particular sampling plan.

Instrument blank - a clean sample processed through the instrumental steps of the measurement process; used to determine the instrument component of the measurement error. See Dynamic blank.

Interference - a positive or negative effect on a measurement caused by a variable other than the one being investigated.

Interference equivalent - the mass or concentration of a foreign substance which gives the same measurement response as one unit of mass or concentration of the substance being measured.

Interlaboratory calibration - the process, procedures, and activities for standardizing a given measurement system to ensure that laboratories participating in the same program can produce comparable data.

Interlaboratory precision - a measure of the variation among laboratories participating in the same test; usually the standard deviation of the data sets.

Interlaboratory test - a test performed by two or more laboratories on the same material for the purpose of assessing the capabilities of an analytical method or for comparing different methods.

Internal quality control - see Intralaboratory quality control.

Internal standard - a standard added to samples in known amount and carried through the procedure as a reference for calibration and controlling instrumental and analytical precision and bias.

Intralaboratory quality control - the routine activities and checks, such as periodic calibrations, duplicate analyses and spiked samples, included in normal internal procedures to control the accuracy and precision of measurements.

Intralaboratory precision - a measure of the method/sample specific analytical variation within a laboratory; usually the standard deviation of a data set. See also Standard deviation and Variance.

Laboratory accreditation - see Accredited laboratory and Accreditation.

Laboratory blank - see Reagent blank.

DRAFT

Laboratory control sample - see Quality control sample.

Laboratory duplicates - two subsamples of a sample analyzed separately with identical procedures. Analysis of laboratory duplicates gives a measure of the precision of laboratory procedures, but not sample collection, preservation or storage internal to the laboratory.

Laboratory performance check solution - a solution of method and surrogate analytes and internal standards; used to evaluate the performance of the instrument system against defined performance criteria.

Laboratory replicates - see Replicate analysis or measurement.

Laboratory spiked blank - see Spiked laboratory blank.

Laboratory spiked sample - see Spiked sample.

Laboratory sample - a subsample of a field, bulk or batch sample selected for laboratory analysis.

Least squares method - a technique for estimating parameters which minimizes the sum of the squares of the differences between each observed value and its corresponding predicted value derived from the assumed model.

Limit of detection (LOD) - see Method detection limit.

Limit of quantification (LOQ) - the concentration of analyte in a specific matrix for which the probability of producing analytical values above the method detection limit is 99 percent.

Linearity - the degree of agreement between the calibration curve of a method and a straight line assumption.

Lot - a number of units of an article or a parcel of articles offered as one item; commonly, one of the units, such as a sample of a substance under study. See Batch.

Lot size - the number of units in a particular lot. See Batch lot and Batch size.

Lower control limit - see Control limit.

Management systems audit - a qualitative on-site evaluation of an organization or laboratory to determine the state of implementation and compliance of its quality assurance program in accordance with its Quality Assurance Program Plan.

Management systems review - a qualitative on-site evaluation of a research or environmental data collection activity that crosses organizational or laboratory lines to determine the state of implementation of each organization's or laboratory's quality assurance program and compliance with the Quality Assurance Program Plan for the data collection activity.

Manual method - an unautomated method.

DRAFT

Matrix - a type of medium (e.g., surface water, drinking water) in which the analyte of interest may be contained. See Medium.

Matrix spike -see Spiked sample.

Matrix spike duplicate sample analysis - see Matrix, Duplicate analysis and Spiked sample.

Maximum contaminant level - the highest permissible concentration of a pollutant that may be delivered to any receptor.

Maximum holding time - the duration for which a sample can be kept under specified conditions without significant degradation of the analyte(s) or property of interest.

Mean - see Arithmetic mean.

Measured value - see Corrected value.

Measurement range - the range over which the precision and/or recovery of a measurement method are regarded as acceptable. See Acceptable quality range.

Measure of central tendency - a statistic that describes the grouping of values in a data set around some common value (e.g., the median, arithmetic mean, or geometric mean.)

Measure of dispersion - a statistic that describes the variation of values in a data set around some common value. See Range, Variance and Standard deviation.

Medium - a substance (e.g., air, water, soil) which serves as a carrier of the analytes of interest. See Matrix.

Medium blank - see Field blank and/or Laboratory blank.

Median - the middle value for an ordered set of n values; represented by the central value when n is odd or by the mean of the two most central values when n is even.

Method - a body of procedures and techniques for performing a task (e.g., sampling, characterization, quantification) systematically presented in the order in which they are to be executed.

Method blank - a clean sample processed simultaneously with and under the same conditions as samples containing an analyte of interest through all steps of the analytical procedure.

Method check sample - see Spiked laboratory blank.

Method detection limit (MDL) - in a given matrix and with a specific method, the minimum concentration of an analyte that can be identified, qualitatively or quantitatively measured, and reported to be greater than zero at the 99% confidence level.

Method of least squares - see Least squares method.

DRAFT

Method quantification limit (MQL) - see Limit of quantification and also Method detection limit.

Minimum detectable level - see Method detection limit.

Mode - the most frequent value or values in a data set.

Multipoint calibration - the determination of correct scale values by measuring or comparing instrument responses at a series of standardized analyte concentrations; used to define the range for generating accurate and precise quantitative data.

Noise - the sum of random errors in the response of a measuring instrument.

Normal distribution - a probability density function that approximates the distribution of random variables and takes the form of the "bell-shaped curve."

Observation - a fact or occurrence that is recognized and recorded.

Observed value - the magnitude of a specific measurement; a variable; a unit of space, time or quantity; a datum. The observed value is that reported before correction for a blank value. See Corrected value.

Outlier - a value that is significantly different, based on an appropriate statistical test, from other members of a data set.

Parameter - any quantity such as a mean or a standard deviation characterizing a population. Commonly misused for "variable", "characteristic" or "property."

Percentage standard deviation - synonym for Coefficient of variation.

Performance evaluation audit - a type of audit in which the quantitative data generated in a measurement system are obtained independently and compared with routinely obtained data to evaluate the proficiency of an analyst or laboratory.

Performance evaluation sample - a sample, the composition of which is unknown to the analyst, provided to test whether the laboratory can produce analytical results within specified limits. See Audit sample, Blind sample and Performance evaluation audit.

Population - all possible items or units which possess a variable of interest and from which samples may be drawn.

Precision - the degree of variation among individual measurements of the same property, usually obtained under similar conditions; a data quality indicator. Precision is usually expressed as standard deviation, variance or range, in either absolute or relative terms.

Preventative maintenance - an orderly program of activities designed to ensure against equipment failure.

Primary reference standard - see Primary standard.

DRAFT

Primary standard - a substance or device, with a property or value that is unquestionably accepted (within specified limits) in establishing the value of the same or related property of another substance or device.

Probability - a numerical estimate of the chance that a particular event will occur. Its value can range from 0 (impossible) up to 1 (certain).

Procedure - a set of systematic instructions for performing an operation.

Proficiency testing - a systematic program in which one or more standardized samples is analyzed by one or more laboratories to determine the capability of each participant.

Property - a quality or trait belonging and peculiar to a thing; a response variable is a measure of a property. Synonym: Characteristic.

Protocol - a detailed written procedure for a field and/or laboratory operation (e.g., sampling, analysis.)

Quality - the sum of features and properties/characteristics of a product or service that bear on its ability to satisfy stated needs.

Quality assurance (QA) - an integrated system of activities involving quality planning, quality control, quality assessment, quality reporting and quality improvement to ensure that a product or service meets defined standards of quality with a stated level of confidence.

Quality Assurance Narrative Statement - a description of the quality assurance and quality control activities to be followed for a research project.

Quality Assurance Objectives - the upper and lower limiting values of the data quality indicators as defined by the data user's acceptable error bounds.

Quality Assurance Program Plan (QAPP) - a formal document describing the management policies, objectives, principles, organizational authority, responsibilities, accountability, and implementation plan of an agency, organization or laboratory for ensuring quality in its products and utility to its users.

Quality Assurance Project Plan (QAPjP) - a formal document describing the detailed quality control procedures by which the data quality requirements defined for the data and decisions in a specific project are to be achieved.

Quality Circle - a small group of individuals from an organization or unit who have related interests and meet regularly to consider problems or other matters related to the quality of the product or process.

Quality control (QC) - the overall system of technical activities whose purpose is to measure and control the quality of a product or service so that it meets the needs of users. The aim is to provide quality that is satisfactory, adequate, dependable, and economical.

DRAFT

Quality control chart - see Control chart.

Quality control check sample - see Calibration check standard.

Quality control sample - a sample prepared from a substance(s) or material(s) of known composition and quantity. It is used to assess the performance of a measurement system or portions thereof. It is intended primarily for routine intralaboratory use in establishing precision and bias. It should be prepared from or be traceable to a standard other than the calibration standard.

Quantitation limits - the maximum or minimum levels or quantities of a target variable that can be quantified with the certainty required by the data user.

Random - lacking a definite plan, purpose or pattern; due to chance.

Random error - that portion of the variance in repeated or replicated sampling and analyses that is due to chance and is not predictable for an individual measurement. The causes of random error are assumed to be indeterminate and/or unassignable and its distribution normal i.e., Gaussian.

Random sample or subsample - a subset of a population or a subset of a sample, selected according to the laws of chance with a randomization procedure.

Randomization - the drawing of samples or subsamples or assigning of material or treatments in a statistically designed experiment or analysis in a test to ensure that initial conditions do not favor one outcome over another.

Randomness - a basic statistical concept and property implying an absence of a plan, purpose or pattern, or of any tendency to favor one outcome rather than another.

Range - the difference between the minimum and the maximum of a set of values.

Raw data - any original factual information from a measurement activity or study recorded in laboratory worksheets, records, memoranda, notes, or exact copies thereof and that are necessary for the reconstruction and evaluation of the report of the activity or study. Raw data may include photographs, microfilm or microfiche copies, computer printouts, magnetic media, including dictated observations, and recorded data from automated instruments. If exact copies of raw data have been prepared (e.g., tapes which have been transcribed verbatim, dated, and verified accurate by signature), the exact copy or exact transcript may be substituted.

Reagent blank - a sample consisting of reagent(s), without the target analyte, introduced into the analytical procedure at the appropriate point and carried through all subsequent steps to determine the contribution to error in the observed value.

Reagent grade - the second highest purity designation for reagents which conform to the current specifications of the American Chemical Society Committee on Analytical Reagents.

DRAFT

Records system (or plan) - a written, documented group of procedures describing required records, steps for producing them, storage conditions, retention period and circumstances for their destruction or other disposition.

Recovery efficiency - in an analytical method, the fraction or percentage of a target analyte extracted from a sample containing a known amount of the analyte.

Reference material - a material or substance, one or more properties of which are sufficiently well established to be used for the calibration of an apparatus, the assessment of a measurement method, or assigning values to materials.

Reference method - a sampling and/or measurement method which has been officially specified by an organization as meeting its data quality requirements.

Reference standard - see Calibration standard.

Relative standard deviation - see Coefficient of variation.

Reliability - the likelihood that an instrument or device will function under defined conditions for a specified period of time.

Repeatability - the capability of a method, test or instrumental procedure to replicate a result under specified conditions (e.g., by a single analyst and/or instrument over a specified period of time.)

Replicability - see Repeatability.

Replicate - an adjective referring to the taking of more than one sample or performance of more than one analysis. Incorrectly used as a noun in place of replicate analysis. Replicate is to be used when referring to more than two items. Duplicate is to be used when there are only two. See Duplicate.

Replicate analysis or measurement - the analysis or measurement of the variable of interest performed as identically as possible on two or more subsamples of a sample. Replicate analyses are used to assess analytical or measurement variance. Often incorrectly called "replicate sample" or "replicate". Replicate analysis is to be used when there are more than two and duplicate analysis if there are only two.

Replicate sample - one of two or more samples taken from the same population and carried through all steps of the sampling and measurement process in an identical manner. Replicate samples are used to assess total (sampling and analysis) method variance. Often incorrectly used in place of replicate analysis.

Representative sample - a sample taken so as to reflect the frequency distribution of the variable of interest in the population as accurately and precisely as specified. To ensure representativeness, the sample may be either completely random or stratified depending upon the conceptualized population and the objective (i.e., upon the decision to be made.)

DRAFT

Representativeness - the degree to which data accurately and precisely represent the frequency distribution of a specific variable in the population; a data quality indicator.

Reproducibility - the extent to which a method, test or experiment yields the same or similar results on repeated trials (e.g., when performed by various analysts or laboratories over an extended period.)

Response variable - see Dependent variable.

Result - the product of a calculation, test method, test or experiment. The result may be a value, data set, statistic, tested hypothesis or an estimated effect.

Risk - the probability or likelihood of an adverse effect.

Robustness - (in)sensitivity of a statistical test method to departures from underlying assumptions. See Ruggedness.

Rounded number - a number, reduced to a specified number of significant digits or decimal places using defined criteria.

Routine method - a defined plan of procedures and techniques used regularly to perform a specific task.

Ruggedness - the (in)sensitivity of an analytical test method to departures from specified analytical or environmental conditions. See Robustness.

Sample - a representative part of a larger whole or a single item of a group; a finite part or subset of a statistical population. A sample serves to provide data or information concerning the properties of the whole group or population.

Sample data custody - see Chain-of-custody.

Sample size - the minimum number of samples needed to furnish sufficient data in a data set to enable the conclusions to be drawn at the required confidence level.

Sampling - the process of obtaining a representative portion of the material of concern from a population or sample.

Sampling equipment blank - a clean sample that is collected in a sample container with the sample-collection device and returned to the laboratory as a sample. Sampling equipment blanks are used to check the cleanliness of sampling devices. See Dynamic blank.

Scheduled maintenance - see Preventative maintenance.

Screening test - a quick test for coarsely assessing a variable of interest.

Secondary standard - a standard whose value is based upon comparison with a primary standard.

DRAFT

Selectivity (analytical chemistry) - the capability of a method or instrument to respond to a target substance or constituent in the presence of nontarget substances.

Sensitivity - capability of method or instrument to discriminate between measurement responses of a variable of interest.

Significant digit - any of the digits 0 through 9, excepting leading zeros and some trailing zeros, which is used with its place value to denote a numerical quantity to a desired rounded number. See Rounded number.

Significant figure - see Significant digit.

Single operator precision - the standard deviation of the results of a series of determinations by the same analyst or operator, all other conditions being equal.

Site - the area within a boundaries established for a defined activity.

Span-drift - the change in the output of a continuous monitoring instrument over a stated time period during which the instrument is not recalibrated.

Span-gas - a gas of known concentration which is used routinely to calibrate the output level of an analyzer. See Calibration check standard.

Specimen - see Sample.

Spike - a known mass of target analyte added to a blank sample or subsample; used to determine recovery efficiency or for other quality control purposes.

Spiked laboratory blank - see Spiked reagent blank.

Spiked reagent blank - a reagent blank fortified with a known mass of the target analyte; usually used to determine the recovery efficiency of the method.

Spiked sample - a sample prepared by adding target or nontarget material to a matrix sample; used, for example, in an analytical method to determine the effect of the matrix on the recovery efficiency.

Spiked sample duplicate analysis - see Duplicate analysis and Spiked sample.

Split sample - one of two or more portions taken from a sample or subsample used to replicate the measurement of the variable(s) of interest.

Standard (measurement) - a substance or material with a property quantified with sufficient accuracy to permit its use to evaluate the same property in a similar substance or material. Standards are generally prepared by placing a reference material in a matrix. See Reference material.

DRAFT

Standard addition - the procedure of adding known increments of the analyte of interest to a sample to cause increases in detection response. The level of the analyte of interest present in the original sample is subsequently established by extrapolation of the plotted responses.

Standard curve - see Calibration curve.

Standard deviation - the most common measure of the dispersion or imprecision of observed values expressed as the positive square root of the variance. See Variance.

Standard material - see Standard (measurement), Reference material.

Standard method - an assemblage of techniques and procedures based on consensus or other criteria, often evaluated for its reliability by collaborative testing and receiving organizational approval.

Standard operating procedure (SOP) - a written document which details the method of an operation, analysis or action whose techniques and procedures are thoroughly prescribed and which is accepted as the method for performing certain routine or repetitive tasks.

Standard reference material (SRM) - a certified reference material produced by the U.S. National Institute of Science and Technology and characterized for absolute content independent of analytical method.

Standard reference sample - see Secondary standard.

Standard solution - a solution containing a known concentration of a standard, prepared and verified by a prescribed method or procedure and used routinely in an analytical method.

Standardization - the process of determining the quantitative relationship between a known mass of target material (e.g., concentration) and the response variable (e.g., the measurement system or instrument response.) See Calibration, Calibration curve and Multipoint calibration.

Statistic - a calculated quantity summarizing the characteristics of a data set (observed or corrected values), e.g., the mean or standard deviation.

Stratification - the division of a target population into groups or strata which are internally homogeneous with respect to the characteristic to be studied.

Stratified sampling - the sampling of internally homogeneous strata. See Stratification.

Stock solution - a concentrated solution of standard(s) or reagent(s) prepared and verified by prescribed procedure(s), and used for preparing working standards.

Subsample - a subset of a sample. A subsample may be taken from any laboratory sample or a field sample. See Aliquant, Aliquot and Test portion.

DRAFT

Surrogate analyte - a pure substance with properties that mimic the analyte of interest. It is unlikely to be found in environmental samples and is added to them for quality control purposes.

Surveillance - the act of maintaining supervision of or vigilance over a well-specified portion of the environment so that detailed information is provided concerning the state of that portion.

Synthetic sample - see Quality control sample.

Systematic error - a consistent deviation in the results of sampling and/or analytical processes from the expected or known value. Such error is caused by human and methodological bias.

Systems error - See Total systems error.

Target - the chosen object of investigation for which qualitative and/or quantitative data or information is desired, e.g., the analyte of interest.

Technical systems audit - a thorough, systematic on-site, qualitative review of facilities, equipment, personnel, training, procedures, recordkeeping, data validation, data management, and reporting aspects of a total measurement system.

Technique - a principle and/or the procedure of its application for performing an operation.

Test - a procedure used to identify or characterize a substance or constituent. See Method.

Test data - see Data.

Test determination - the process of obtaining an observed value of a property for a single sample. Also, the observed value obtained (i.e., the test result.) See Test result.

Test method - see Method.

Test portion - a subsample of the proper amount for analysis and measurement of the property of interest. A test portion may be taken from the bulk sample directly, but often preliminary operations, such as mixing or further reduction in particle size, are necessary. See Subsample.

Test result - a product obtained from performing a test determination. See Test determination.

Test sample - see Test portion.

Test specimen - see Test portion.

Test unit - see Test portion.

DRAFT

Time-proportioned sample - a sample composed of samples collected at preselected uniform time intervals.

Total Quality Management (TQM) - the process whereby an entire organization, led by senior management, commits to focusing on quality as a first priority in every activity. TQM implementation creates a culture in which everyone in the organization shares the responsibility for continuously improving the quality of products and services, and for "doing the right thing, the right way, the first time, on time."

Total systems error - the sum of statistical variation in test results from sampling field and laboratory activities.

Traceability - an unbroken trail of accountability for verifying or validating the chain-of-custody of samples, data, the documentation of a procedure, or the values of a standard.

Treatment (experimental) - an experimental procedure whose effect is to be measured and compared with other treatments.

Trip blank - a clean sample of matrix that is carried to the sampling site and transported to the laboratory for analysis without having been exposed to sampling procedures.

Tuning - the process of adjusting a measurement device or instrument, prior to its use, to ensure that it works properly and meets established performance criteria.

Type I error, (alpha error) - an (incorrect) decision resulting from the rejection of a true hypothesis. (A false positive decision.)

Type II error, (beta error) - an (incorrect) decision resulting from acceptance of a false hypothesis. (A false negative decision.)

Uncertainty - a measure of the total variability associated with sampling and measuring, taking into account two major error components: systematic error (bias) and random error.

Universe - see Population.

Upper control limit - see Control limit.

User check - an evaluation of a written procedure for clarity and accuracy in which an independent laboratory analyzes a small number of spiked samples, following the procedure exactly.

Valid study - a study conducted in accordance with accepted scientific methodology, the results of which satisfy pre-defined criteria.

DRAFT

Validated method - a method which has been determined to meet certain performance criteria for sampling and/or measurement operations.

Validation - the process of substantiating specified performance criteria.

Value - the magnitude of a quantity. A single piece of factual information obtained by observation or measurement and used as a basis of calculation.

Variable - an entity subject to variation or change.

Variance (statistical definition) - a measure of the dispersion of a set. The sum of the squares of the difference between the individual values of a set and the arithmetic mean of the set, divided by one less than the number of values in the set. (The square of the standard deviation.) See also Measure of dispersion.

Verifiable - the ability to be proven or substantiated.

Warning limit - a specified boundary on a control chart that indicates a process may be going out of statistical control and that certain precautions are required. The boundary specified is usually two standard deviations (i.e.,the 95% confidence level.)

Within-laboratory standard deviation - see Intralaboratory precision.

Working standard - see Secondary standard.

Zero drift - the change in instrument output over a stated time period of nonrecalibrated, continuous operation, when the initial input concentration is zero; usually expressed as a percentage of the full scale response.

APPENDIX C

List of Participants

Agemian, Haig
National Water Quality Laboratory
Environment Canada
Burlington ON

Arafat, Nabil
National Water Research Institute
Environment Canada
Burlington ON

Aspila, Keijo
National Water Research Institute
Environment Canada
Burlington ON

Balsinger, Jack
Midwest Research Institute
Kansas City MO U.S.A.

Barbour, Michael
EA Engineering, Science, and Technology
Sparks MD U.S.A.

Bigelow, David
Grasslands Laboratory
Colorado State University
Fort Collins CO U.S.A.

Bisson, Robert
National Water Research Institute
Environment Canada
Burlington ON

Breckenridge, Robert
Environmental Research
Idaho National Engineering Laboratory
Idaho Falls ID U.S.A.

Burkman, Bill
NSI Technology Services Corporation
Radnor PA U.S.A.

Chapman, Peter
E.V.S. Consultants
North Vancouver BC

Chau, Alfie
National Water Research Institute
Environment Canada
Burlington ON

Cline, Steve
NSI Technology Services Corporation
Corvallis OR U.S.A.

Coffey, Deborah
NSI Technology Services Corporation
Corvallis OR U.S.A.

Comba, Michael
National Water Research Institute
Environment Canada
Burlington ON

Craig, Gordon
Beak Consultants Limited
Brampton ON

Cumming, Brian
Department of Biology
Queen's University
Kingston ON

Davies, Ian
Freshwater Institute
Fisheries and Oceans Canada
Winnipeg MB

Day, Kristin
National Water Research Institute
Environment Canada
Burlington ON

Doull, James
Water Quality Branch
Environment Canada
Moncton NB

Dufresne, Alain
Department of Renewable Resources
Macdonald College
Sainte-Anne-de-Bellevue PQ

Duval, Dominique
St. Lawrence Centre Laboratory
Environment Canada
Longueuil PQ

Dwire, Kate
NSI Technology Services Corporation
Corvallis OR U.S.A.

Ellingson, Lisa
NSI Technology Services Corporation
Corvallis OR U.S.A.

Ervin, Valerie
HJF Forestry Centre
Forestry Canada
Fredericton NB

Forbes, Michael
National Water Quality Laboratory
Environment Canada
Burlington ON

Foster, Brian
Laboratory Services Branch
Ontario Ministry of the Environment
Rexdale ON

Gaskin, James
Water Quality Branch
Inland Waters Directorate
Environment Canada
Hull PQ

Graves, Robert
U.S. Environmental Protection Agency
Cincinnati OH U.S.A.

Hart, Don
Beak Consultants Limited
Brampton ON

Innes, John
Forestry Commission
Farnham Surrey UK

Jamro, Ghulam
National Water Quality Laboratory
Environment Canada
Burlington ON

Johannessen, Merete
Norwegian Institute for Water Research (NIVA)
N-0808 Oslo 8 NOR

Kinch, Craig
Beak Consultants Limited
Brampton ON

King, Donald E.
Laboratory Services Branch
Ontario Ministry of the Environment
Rexdale ON

Kingston, John
Department of Biology
Queen's University
Kingston ON

Kirkland, Linda
U.S. Environmental Protection Agency
Washington DC U.S.A.

Laflamme, Denis
Dir. recherche et développement
Ministere de l'Energie et des Ressources
Sainte-Foy PQ

Lalonde, Hélène
Department of Renewable Resources
Macdonald College
Sainte-Anne-de-Bellevue PQ

Lawrence, John
National Water Research Institute
Environment Canada
Burlington ON

Lazorchak, James
Environmental Monitoring Systems Lab.
U.S. Environmental Protection Agency
Cincinnati OH U.S.A.

Le Page, Alain
Forestry Canada - Quebec Region
Sainte-Foy PQ

Linsey, Garry
Freshwater Institute
Fisheries and Oceans Canada
Winnipeg MB

Manen, Carol-Ann
National Oceanic and Atmospheric Administration
National Ocean Service
U.S. Department of Commerce
Rockville MD U.S.A.

Marsh, Marius
Air Resources Branch
Ontario Ministry of the Environment
Toronto ON

McKenzie, Dan
Environmental Research Laboratory
U.S. Environmental Protection Agency
Corvallis OR U.S.A.

McMullen, Dennis
Technology Applications, Inc.
Cincinnati OH U.S.A.

McQuaker, Neil
B.C. Ministry of the Environment
Vancouver BC

Michael, Dan
Research Triangle Institute
Washington DC U.S.A.

Mickler, Robert
U.S. Forest Service
Southeastern forest experiment station
Research Triangle Park NC U.S.A.

Millers, Imants
NA State and Private Forestry
US Forest Service
Durham NH U.S.A.

Morrison, Ian
Great Lakes Forestry Centre
Forestry Canada
Sault Ste. Marie ON

Neptune, Dean
Quality Assurance Management Staff
U.S. Environmental Protection Agency
Washington DC U.S.A.

Palmer, Craig
NSI Technology Services Corporation
Corvallis OR U.S.A.

Paul, John
U.S. Environmental Protection Agency
Narragansett RI U.S.A.

Peck, Dave
Lockheed Engineering & Sciences Co.
Las Vegas NV U.S.A.

Philbert, Francis
Inland Waters Directorate-Ontario Region
Environment Canada
Burlington ON

Pollard, Jim
Lockheed Engineering & Sciences Co.
Las Vegas NV U.S.A.

Pollock, Thomas
Water Quality Branch
Environment Canada
Moncton NB

Reid, Neville
Air Resources Branch
Ontario Ministry of the Environment
Toronto ON

Samant, Hari
Environmental Protection - Atl. Region
Environment Canada
Dartmouth NS

Samoiloff, Martin
Bioquest International Inc.
Winnipeg MB

Schumacher, Brian
Lockheed Engineering & Sciences Co.
Las Vegas NV U.S.A.

Sergeant, David
Great Lakes Laboratory for Fisheries and Aquatic Sciences
Fisheries and Oceans Canada
Burlington ON

Stainton, Mike
Freshwater Institute
Fisheries and Oceans Canada
Winnipeg MB

Strachan, Bill
National Water Research Institute
Environment Canada
Burlington ON

Thornley, Stew
Ontario Ministry of the Environment
London ON

Turle, Richard
Wildlife Toxicology and Survey Branch
Canadian Wildlife Service
Environment Canada
Hull PQ

Van Remortel, Rick
Lockheed Engineering & Sciences Co.
Las Vegas NV U.S.A.

Vereault II, Donald
Environmental Chemistry Lab
Sawyer Environmental Research Centre
University of Maine
Orono ME U.S.A.

Villard, Serge
Laboratory Services Branch
Ontario Ministry of the Environment
Rexdale ON

Walsh, Michael
Water Resources Branch
Ontario Ministry of the Environment
Toronto ON

Watt, Evan
Water Quality Branch
Inland Waters Directorate
Environment Canada
Hull PQ

Winter, Helmut
Environnement Quebec
Sainte-Foy PQ

Wismer, Don
Environmental Studies & Assessment Dept.
Ontario Hydro
Toronto ON

Wotton, David
Manitoba Environment
Winnipeg MBM